Improving Salience-based Multi-document Summarization Performance Using a Hybrid Sentence Similarity Measure

Kamal Sarkar and Sohini Roy Chowdhury

[1] Jadavpur University, 188, Raja S.C. Mallick Rd, Kolkata-700032,West Bengal,India
[2] RCC Institute Of Information Technology, Canal South Rd, Beleghata, Kolkata 700015,West Bengal, India

Abstract. The process of creating a single summary from a group of related text documents obtained from many sources is known as multi-document summarization. The efficacy of a multi-document summarization system is heavily reliant upon the sentence similarity metric employed to eliminate redundant sentences from the summary, given that the documents may contain redundant information. The sentence similarity measure is also crucial for a graph-based multi-document summarization, where the presence of an edge between two phrases is decided by how similar the two sentences are to one another. To enhance multi-document summarization performance, this study provides a new method for defining a hybrid sentence similarity measure combining a lexical similarity measure and a BERT-based semantic similarity measure. Tests conducted on the benchmark datasets demonstrate how well the proposed hybrid sentence similarity metric is effective for enhancing multi-document summarization performance.

Keywords: Extractive Summarization. Multi-Document Text Summarization. BERT. Hybrid Similarity measure. Semantic Similarity similarity, Lexical similarity

1 Introduction

Text summarization involves extracting vital information from a given text or a given document set of related text documents while preserving the fundamental concepts and primary ideas in the produced summary. The goal is to create a concise and coherent summary that captures the essence of the original text. It addresses the challenge of information overload by providing a more digestible form of content, saving time, and aiding in efficient information retrieval and document understanding across various applications.

Multi-document summarization is a task in natural language processing (NLP) that involves generating a concise and coherent summary from multiple documents. This can be particularly useful when dealing with a large amount of information or when trying to distill key information from a variety of sources. It is an evolving field, and ongoing research aims to address its challenges and enhance its effectiveness across various applications. The extractive summarization method involves

David C. Wyld et al. (Eds): IOTBS, NLTM, AIMLA, DBDM -2024
pp. 15-29, 2024.CS & IT -CSCP 2024 DOI: 10.5121/csit.2024.140202

extracting existing sentences or phrases directly from the source documents, rather than creating entirely new content. It depends on identifying and selecting the most crucial sentences based on criteria like importance, relevance, or frequency. The objective of multi-document extractive text summarization is to condense the vital information and primary ideas from a set of documents related to a topic or event, offering a succinct representation of the collective content. This method proves valuable when handling a substantial amount of information from various sources, enabling users to manage the information overload problem by swiftly understanding the key points without having to review the entire document set.

There are many earliest approaches in the field of multidocument extractive text summarization. A document extraction strategy for multi-document summarization that extends single-document summarization techniques by incorporating supplementary information regarding the entire document set and the interconnections between the documents was proposed by [1]. A graph-based approach was developed by [2] which considered a matrix of connections weights derived from cosine similarity between sentences (nodes) and used centrality-based salience for creating a summary from multiple documents. Radev et. al. used [2] a modified Cosine similarity measure which is based on the geometric interpretation of sentences into vectors and does not consider the semantic meaning of words. Therefore, two sentences with similar meanings but different word choices may have a lower cosine similarity. It cannot capture more complex semantic relationships between sentences. It treats each term independently and doesn't consider the contextual meaning or relationships between terms. If two sentences have dissimilar terms but they share one highly frequent term common between them, the similarity value might be high. If sentence lengths vary significantly, cosine similarity may not adequately normalize for this. Longer sentences may have inherently lower cosine similarity scores, leading to biased results.

The methodology [2] [35] [28] [36] [37] employed for text summarization involves the use of graphs. In this approach, sentences within a document or document set are depicted as nodes in a graph. The connections between pairs of sentences are established based on the degree of similarity between them. Assessing the significance of a sentence is done through a graph-based method that relies on both global and local information from the entire graph. Additionally, graph-based methods predominantly employ the standard cosine similarity measure to construct the similarity graph. Numerous existing extractive summarization systems mentioned earlier employ sentence similarity to either reduce redundancy, construct a graph, or both. Consequently, we propose a hypothesis that suggests enhancing the similarity measure can lead to an improvement in the performance of graph-based summarization. Our paper is arranged in the following way. Related work is discussed in section 2. The methodology is explained in section 3. Section 4 highlights the evaluation metric and results. Section 5 concludes the paper.

2 Related work

Different approaches for performing extractive multi-document text summarization have been proposed over time. The initial studies on extractive summarization involve ranking sentences using basic features like sentence position, term frequency, or specific key phrases [4–8]. The next step is to select the top n non-redundant sentences based on the compression ratio. In [9], a method based on information extraction is introduced for multi-document summarization. This method identifies similarities and differences across the documents in the set. An improved technique for calculating sentence similarity with the goal of enhancing the performance of multi-document summarization was proposed by [10]. The prevalent approach in automated extractive summarization involves scoring phrases or sentences to generate summaries. Sentence scoring is widely embraced in the majority of contemporary methods. Scoring techniques are categorized into word scoring, sentence scoring, and graph scoring [11]. In word scoring techniques, sentences are assigned scores based on the significance of words and their frequency in the text [6] ,[12], [13]. Notably, words like proper nouns, places, and objects, considered determinants, receive higher scores [14, 15]. Text scoring methods consider formal properties such as bold, italicized, and underlined words [16]. Sentences beginning with phrases like 'Briefly,' 'Finally,' and 'As a result' are identified as sign phrases and subsequent sentences are deemed important [12]. Similarly, evaluation involves the text title, with sentences containing title words considered for inclusion in the summary, increasing their importance [17]. Sentence scoring methods also consider sentence length, giving more weight to longer sentences [16], [18]. Scoring involves assigning points based on sentence position and whether it includes numerical values [12],[14],[19]. In Reference [20], the authors outlined an approach for extractive summarization designed to aid learners facing reading difficulties. Graph-based representations are commonly employed in text analysis methodologies due to their highly effective solutions. Reference [28] introduced TextRank, incorporating a graph-based representation to summarize text by identifying intersections in the content. Similarly, LexRank, presented in Reference [2], utilizes an eigenvector centrality-based algorithm, a form of node centrality method. Both TextRank and LexRank draw inspiration from the PageRank algorithm [21], a framework for document summarization that identifies central sentences based on mutual information between term and sentence sets [22].

Random Walk has been employed to generate summaries of primary documents. In Reference [27], a summarization system targeting the biomedical domain was introduced. Utilizing the Unified Medical Language system, a graph was derived from concepts and relationships using a semi-dictionary-based approach, followed by the application of the PageRank algorithm. In Reference [29], a novel graph based on reinforced random walk was suggested. In Reference [23], a multi-layered representation involving documents, sentences, and words was utilized. The authors in

Reference [24] incorporated graphs to depict documents, employing link generation for automated document summarization. They established the document structure by revealing text relationships and evaluating summaries through comparison with manually created ones. Reference [25] introduced a graph-based approach to ensure semantic continuity, where nodes represented document terms, and edges reflected semantic relationships. The graph diameter calculation for all nodes described the shortest and longest paths as the weakest and strongest bonds. In Reference [26], graph structures and documents were defined, and nodes and edges were established based on local similarities.

In our study, we have computed the summary worthiness of a sentence considering its salience which is measured based on the similarity of the sentence to other sentences in the input. Our primary contribution is to define the similarity measure by linearly combining a TF-IDF-base lexical similarity measure and a BERT-based similarity measure. The overall score of a sentence is computed by combining the salience-based score and sentence positional score.

3 Proposed Method

Erkan et. al. proposed in [2] a graph-based degree centrality for computing sentence importance. In this approach, the input document set is converted into a collection of sentences, and the sentence collection is represented as a graph where each node corresponds to a sentence and an edge between any two nodes is established if the lexical cosine similarity between the corresponding two sentences crosses a predefined cut-off value. In the degree centrality-based approach, the degree of a node is considered as the salience score of the corresponding sentence. During summary generation, the sentences are ranked based on their salience scores, and the top sentence is selected first in the summary. The next sentence is chosen from the ranked list and it is included in the summary if it is sufficiently not similar to the previously selected sentences. During summary generation, the sentence comparison is done using a lexical similarity measure for redundancy removal.

One of the main drawbacks of the approach proposed by Erkan et. al.[2] is that it does not use semantic similarity measures in graph construction as well as redundancy removal. To overcome this drawback, we propose a degree centrality method that uses a hybrid sentence similarity measure that combines the lexical sentence similarity used by Erkan et. al.[2] and BERT-based semantic similarity. The proposed hybrid sentence similarity measure is used in graph construction as well as redundancy removal. The redundancy plays a crucial role in multi-document summarization. In the subsequent subsections, we will present the first two similarity measures: Lexical and semantic.

3.1 Lexical Similarity

In this method, a sentence is represented as a TF-IDF vector whose length is equal to the vocabulary size, and the similarity between two sentences is computed as the cosine of the vectors for the corresponding two sentences. We call it lexical similarity because it breaks sentences into the bags of words. The similarity between two sentences $S1$ and $S2$, represented by vector representations using $TF * IDF$ values, is calculated using the cosine similarity formula [2]:

$$LexSim(s_1, s_2) = \frac{\sum_{j=1}^{n} w_{1j} \cdot w_{2j}}{\sqrt{\sum_{j=1}^{n} (w_{1j})^2} \cdot \sqrt{\sum_{j=1}^{n} (w_{2j})^2}} \tag{1}$$

where S_1=$(w_{11}, w_{12}, w_{13}.........w_{1n})$ and
S_2=$(w_{21}, w_{22}, w_{23}.........w_{2n})$ which are TF-IDF based vector representation pf the sentences.
$w_{ij} = TF_{ij} * IDF_{ij}$
The $TF * IDF$ value for a word, denoted as w_{ij}, is computed as the product of the Term Frequency (TF) and the Inverse Document Frequency (IDF). The TF is calculated using formula [2] [36]:

$$TF_{ij} = n(i, j) \tag{2}$$

Where $n(i, j)$ is the number of occurrences of the word i in sentence j.
The IDF is calculated as:

$$IDF_i = \log\left(\frac{N}{n_i}\right) \tag{3}$$

Where N is the total number of documents in a corpus, and n_i is the number of documents containing the word i. We have set a threshold value of 3 for filtering out the noisy words. The Words whose TF*IDF value is >3 are considered while computing a sentence vector. Here this similarity is noted as $LexSim$.

3.2 BERT based semantic similarity measure

BERT(Bidirectional Encoder Representations from Transformers) is based on the transformer architecture, which was introduced by [3]. Transformers have become a foundational architecture for various NLP tasks due to their effectiveness in capturing long-range dependencies in sequences. Unlike earlier sequence models processing text in one direction, BERT employs bidirectional processing, taking into account context from both left and right directions. We have obtained a sentence vector using the BERT encoder. The sentence vector obtained by the BERT encoder is 768 dimensional. The similarity between two sentences is computed using the Cosine of the corresponding BERT vectors. Equation 4 is used to compute BERT-based semantic similarity.

$$SemanticSim_{bert} = cosine(SVi_{bert}, SVj_{bert}) \quad (4)$$

In the above equation, SVi_bert and SVj_bert are two 768-dimensional vectors obtained from the BERT encoder for sentences i and j.

Equation 4 is a semantic similarity measure because we use a BERT encoder for sentence representation which is a mapping of a sentence to a high-level abstract space

3.3 Hybrid Similarity Measure

Although the lexical similarity measure given in Equation 1 takes into account the relative importance of the terms in the input, it considers the bag-of-words model for sentence representation. Hence the sentence vector becomes sparse. On the other hand, the semantic similarity measure given in 1 finds how much two sentences are contextually and semantically similar. In some cases, it is observed that the semantic similarity output is very high although human finds that the concerned sentences have low semantic similarity. It may happen when the sentences do not contain sufficient information or one sentence is long and another sentence is short.
To generate a more precise measure of sentence similarity, we hybridize both similarity metrics - 1) lexical similarity and 2) BERT-based semantic similarity. We have performed hybridization using a blending parameter α shown in Equation 5. α is tuned to find the appropriate weights for the similarity measures which are combined.

$$Sim_{ij} = \alpha * Lexsim(i, j) * (1 - \alpha) * SemanticSim_{bert}(i, j) \quad (5)$$

in Equation 5, Sim_{ij} = Similarity among i_{th} sentence and j_{th}sentence.
α=Blending parameter

3.4 Summarization method

The proposed summarization system undergoes several key stages: Pre-processing, Graph formation, Centrality computation, and summary creation.

Pre-processing In this phase, sentences are delimited using a sentence tokenizer of the NLTK toolkit. Then sentences are broken into a collection of words. We have removed stop words from the dataset. Stop words are common words that frequently occur in the dataset but are unimportant. NLTK toolkit was used to remove stop words from sentences. Figure 1 shows a sample sentence after the removal of stop words.

Sample sentence:
The impact analysis of major events like the Covid-19 pandemic is fundamental to research in social sciences
After removing stop word from sentence:
impact analysis major events like Covid-19 pandemic fundamental research social sciences

Fig. 1. Removal of stop word

Graph Construction and Centrality Calculation To compute degree centrality, we need to construct a graph wherein each sentence corresponds to a node and establish an edge between two nodes if the hybrid similarity between the corresponding sentences is greater than a threshold. In our experiment, we get the best result by setting the threshold value to 0.6. We consider this threshold value since we take notable similarities.

The graph is represented as an adjacency matrix which we call the sentence similarity matrix in which each cell (i,j) contains the value of the hybrid similarity (equation5) between sentence i and sentence j. Figure 2 illustrates a similarity graph with the pairs of sentences with similarities exceeding 0.1, 0.2, and 0.3 respectively. In the graph, $D_i s_j$ refers to the j-th sentence in the i-th document, and the different types of lines connecting nodes are used to indicate varying similarity values. The degree centrality of a sentence is defined as the degree of the corresponding node within the similarity graph. The degree centrality score of a sentence is regarded as the degree of the corresponding node in the similarity graph. Given that centrality represents the number of edges incident on a node, these scores vary from 1 to the total number of sentences in the document set. Consequently, normalization is necessary to constrain this value within a range of 0 to 1. To achieve this, we employ the min-max procedure as follows.

$$c_{score} = \frac{d - min}{max - min} \tag{6}$$

Where: c_{score} represents the normalized centrality score of the sentence, d denotes the degree of the sentence representing a node in the similarity graph), min stands for the minimum degree value in the graph, and max signifies the maximum degree value in the graph.

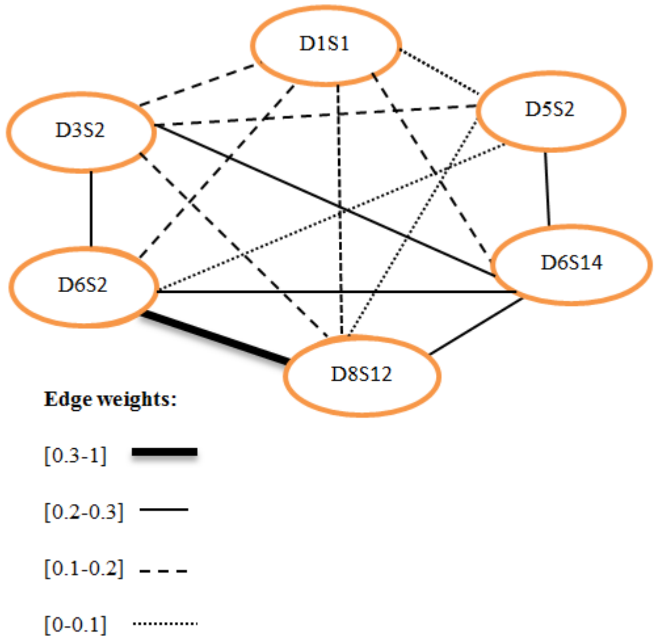

Fig. 2. Weighted similarity graph

3.5 Sentence ranking and summary generation

To generate the summary from a document set, the sentences are ranked based on their scores where the score of a sentence is a linear combination of the normalized centrality score and the positional score of the sentence. We consider positional score because it is proven to be an effective feature in text summarization [9] [38].

The overall score of each sentence is derived through the linear combination of centrality score (c_{score}) and positional score. The Positional score is calculated using Equation7. Sentences are ranked in decreasing order of their combined scores. To generate the non-redundant summary, the top-ranked sentence is selected first and then the next sentence from the ranked list is selected in the summary if it is sufficiently dissimilar to the sentences already selected in the summary. This process is continued until the desired summary length is reached. We use a similarity threshold value for deciding whether two sentences are similar or not. This threshold value is also tuned to find its optimal value.

$$positional - score = \frac{1}{\sqrt{i}} \tag{7}$$

where i is the position of a sentence in the document

4 Evaluation and Results

4.1 Evaluation metric:

We have evaluated our approach using an automatic summary evaluation package ROUGE which is widely used by many researchers. ROUGE [30] measures n-gram overlap between a system-generated summary and the reference summaries [31]. ROUGE counts various kinds of overlapping units between the system summary and the reference summaries. We have used the latest version of the ROUGE package - ROUGE 1.5.5 for evaluating the system summaries. The ROUGE toolkit reports various ROUGE–N scores, for example, ROUGE-1, ROUGE-2, etc. Along with ROUGE-1 scores, many state-of-the-art summarization systems have been evaluated using ROUGE-2 (bigram-based), and ROUGE-SU4 (skip bigrams with skip distance up to 4 words [30]. So, we consider ROUGE-1, ROUGE-2, and ROUGE-SU4 scores for evaluating our proposed summarization models. We set the summary length to 665 bytes(100 words) as per DUC 2004 guidelines since we tested the proposed model on the DUC 2004 dataset. We use ROUGE-F score scores to evaluate and compare our proposed summarization method with other existing summarization methods.

4.2 Results:

We have used the DUC2004 dataset[3] for the proposed summarization model evaluation. The data set is comprised of 50 folders. Each folder contains approximately 10 documents which are news articles sourced from both the New York Times and Associated Press Wire services. Each article is accompanied by four distinct summaries, The system summaries are compared with reference summaries using the ROUGE package for evaluating the model's performance.

Since the blending parameter α was used to find the weights for the similarity measures which are combined to have a hybrid similarity measure, it has an impact on the summarization output. To find the optimal value of α, We have varied the blending parameter α to obtain the best performance of the proposed model. The results with the different values of α are shown in figure 3.

As we can see from Figure 3, the best result is obtained when the value of α is set to 0.8. To generate the non-redundant summary, we consider a similarity threshold as mentioned in the summary generation subjection. To assess the influence of this similarity threshold value used for minimizing redundancy in the summary, we have varied it while keeping the blending parameter α fixed to the optimal value of 0.8. As shown in figure 4, the best result is obtained when we set the similarity

[3] https://duc.nist.gov/duc2004/

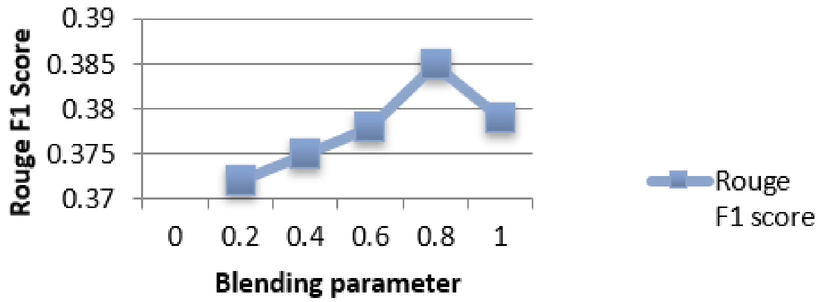

Fig. 3. Impact on summarization performance while varying the values of the blending parameter, α

Table 1. The performance of the proposed summarization method(degree + position) with hybrid similarity measure on DUC2004 data. R-1 implies Rouge-1 F1 score, R-L implies Rouge-L F1 score and R-SU4 indicates Rouge-SU4 F1 score

Alpha(α)	**R-1**	**R-L**	**R-Su4**
0.4	0.3756	0.3461	0.1645
0.6	0.3789	0.3477	0.1659
0.8	**0.3855**	**0.3591**	**0.1797**

threshold to 0.6.

We have shown the performance of the proposed model in Table 1. Since the proposed work is an extension of the degree centrality method proposed by Erkan et. al (2004) [2] who did not use the positional feature in sentence ranking, we have evaluated the proposed method without positional information and the obtained results are given in Table 2. In both the tables, the best scores are indicated by bold font.

Table 2. The performance of the proposed summarization method (degree only) on DUC2004 data. R-1 implies Rouge-1 F1 score, R-L implies Rouge-L F1 score and R-SU4 indicates Rouge-SU4 F1 score

Alpha(α)	**R-1**	**R-L**	**R-Su4**
0.4	0.3607	0.3302	0.1459
0.6	0.3619	0.3316	0.1463
0.8	**0.3719**	**0.3348**	**0.1649**

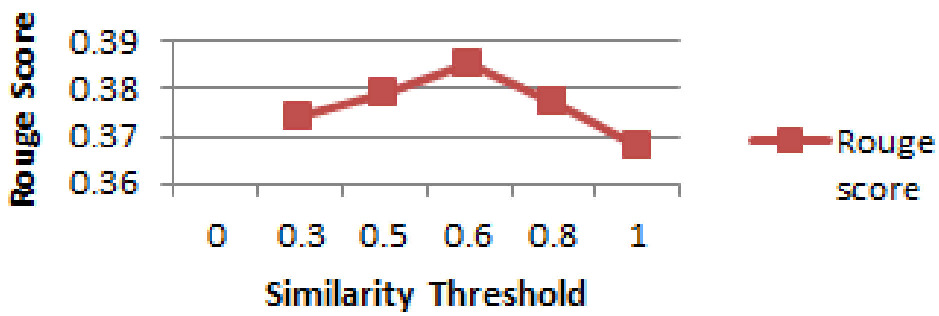

Fig. 4. Similarity threshold vs Rouge F1 score

4.3 Comparisons with Existing methods

Since the proposed work is an extension of the degree centrality method proposed by Erkan et. al (2004) [2] who did not use the positional feature in sentence ranking, we have evaluated our proposed method without positional information and compared the obtained results with that of the degree centrality method proposed by Erkan et. al. [2]. In Table 3, for comparison, we have presented the results obtained by three models- the proposed method with only degree centrality score, the proposed method (degree centrality score + positional score), and the degree centrality method with lexical similarity as proposed in the paper of Erkan et. al [2]. Since we have used the ROUGE F1 score for system evaluation whereas Erkan et. al used the ROUGE-1 recall score for system evaluation, we have implemented the degree centrality method with the lexical similarity measure as proposed in [2]. In table 3, we have shown the comparison of the proposed method with the degree centrality-based method proposed by Erkan et. al. It is evident from table 3 that our proposed method with degree centrality feature performs significantly better than that of the degree centrality-based method proposed by Erkan et. al. who used only lexical similarity for constructing graph and redundancy removal whereas we used a hybrid similarity measure. It shows that the hybrid similarity measure is effective. The table also shows that the performance of the proposed method (degree + position) improves 3.65% (=(((0.3855- 0.3719)/0.3719)*100)) over the proposed method with the degree feature only.

We conducted a comparison of the proposed summarization method with those that participated in DUC 2004 Task 2. The ROUGE scores for the systems that participated in the DUC 2004 Task 2 have been taken from the paper of Sarkar et. al. (2015) [10]. The summaries generated by the participating teams during the DUC 2004 contest were made available by the conference authority on their website [4]. The leading teams that took part in DUC 2004 are identified by peer

[4] http://duc.nist.gov

Table 3. Comparison with existing models that used Degree centrality score only.

Model	Rouge-1 F1 score
Our proposed model (degree + position)	0.3855
Our proposed model(degree centrality score only)	0.3719
The Degree centrality based approach proposed in [2]	0.3685

codes 65, 104, and 35. Among these, the team assigned peer code 65 emerged as the top-performing system in DUC 2004. Comparisons of the proposed model with the systems that participated in the DUC 2004 Task 2 are shown in Table 4. in this table, We have also compared the proposed method with three existing multi-document summarization systems (MDS)- two systems proposed by Sarkar et. al in 2015 [10] and 2022 [32], and another centroid-based system called Mead developed by Radev et. al. [2].
A short description of these three existing models is as follows.

– The method proposed by Sarkar et. al in 2015 [10] is most similar to our proposed method because they used degree centrality and positional information. However, our method differs in the similarity measure used for constructing a graph and redundancy removal. Sarkar et. al(2015) used a hybrid similarity measure that combines two lexical similarity measures whereas our method uses a hybrid similarity measure that combines a lexical similarity measure and a BERT-based semantic similarity measure.
– The method proposed by Sarkar et. al in 2022 [32] used semantic term relations for finding the term weights. The sentence score is a linear combination of the term weight-based score and the positional score. To find term relations, they used cosine similarity between the word embedding-based representation vectors for the terms.
– Method developed by Radev et. al. [2] was incorporated in an MDS system called Mead that combines centroid score with positional feature and sentence length feature. In this work, the terms whose TF*IDF weights were greater than a predefined threshold value were considered centroid. The similarity of a sentence with the centroid is taken as the sentence score which is further combined with the positional score for finding the overall sentence score.

As we can see from Table 4, the proposed method performs better than all existing methods presented in the table.

Table 4. Comparison with the systems that participated in the DUC 2004 Task 2 and three existing MDS systems

Model	Rouge-1 F1 score
Our proposed method(degree + position) with hybrid similarity ($\alpha = 0.8$)	**0.3855**
The method proposed in [32]	0.3840
A Graph-based system with a hybrid similarity measure proposed in [10]	0.3820
DUC Sys65	0.3795
DUC Sys35	0.3757
MEAD baseline[2]	0.3737
Sys104	0.3712
DUC Coverage baseline	0.3454
DUC Lead baseline	0.3210

5 Conclusion

In this paper, we have proposed a hybrid similarity measure for performing extractive multi-document text summarization. The proposed hybrid similarity measure combines a lexical similarity measure with a BERT-based semantic similarity measure. The proposed hybrid similarity measure has been used to construct a graph and redundancy removal. The degree centrality and positional features have been used for sentence ranking. The experimental results reveal that the proposed method with only the degree centrality score performs better than the degree centrality-based method with lexical similarity measure. It proves that the amalgamation of the BERT-based semantic similarity measure with the lexical similarity is effective in improving salience-based multi-document text summarization.
Although the proposed method has achieved satisfactory results, there is a scope for improvement. We have used the BERTbase model for sentence representation. In the future, we would like to use the BERTlarge model or other large language models for sentence representation.

References

1. J. Goldstein et al. Multi-Document Summarization By Sentence Extraction. NAACL-ANLP 2000 Workshop on Automatic Summarization,2002.
2. G. Erkan and D. R. Radev. LexRank: graph-based lexical centrality as salience in text summarization. J. Artif. Int. Res. 22, 1 (July 2004), 457–479.
3. A. Vaswani, et al. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010, 2017.
4. P. B. Baxendale, "Man-made index for technical literature—An experiment," IBM Journal of Research and Development, 2(4), pp. 354–361, 1958.

5. H. P. Edmundson, “New methods in automatic extracting,” Journal of the Association for Computing Machinery, 16(2), pp. 264–285, 1969.
6. H. P Luhn, “The automatic creation of literature abstracts,” IBM Journal of Research Development, 2(2), pp. 159–165, 1958.
7. K. Sarkar, “Using domain knowledge for text summarization in medical domain,” International Journal of Recent Trends in Engineering, 1(1), pp. 200-205, 2009.
8. K. Sarkar, M. Nasipuri, and S. Ghose, “Using machine learning for medical document summarization,” International Journal of Database Theory and Application, vol. 4, pp. 31-49, 2011.
9. K.R. McKeown, and D. R. Radev, “ Generating summaries of multiple news articles,” In Proceedings of the 18th Annual International ACM SIGIR Conference on Re-search and Development in Information Retrieval, Seattle, pp. 74–82, 1995.
10. K. Sarkar, K. Saraf and A. Ghosh, ”Improving graph based multidocument text summarization using an enhanced sentence similarity measure,” 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), Kolkata, India, 2015, pp. 359-365.
11. Ferreira R et al. Assessing sentence scoring techniques for extractive text summarization, 2013.
12. R. Shardan. U.V. Kulkarni. Implementation and evaluation of evolutionary connectionist approaches to automated text summarization, 2010.
13. M.N Azadani, N. Ghadiri, E. Davoodijam. Graph-based biomedical text summarization: an itemset mining and sentence clustering approach. J Biomed Inform 2018;84:42–58.
14. D. Kadam “A Comparative Study Of Hindi Text Summarization Techniques: Genetic Algorithm And Neural Network.” (2015).
15. M. Gambhir and V. Gupta Recent automatic text summarization techniques: a survey. Artif Intell Rev 2017;47(1):1–66.
16. V.Gupta. Hybrid algorithm for multilingual summarization of Hindi and Punjabi documents. In: Mining intelligence and knowledge exploration. Springer; 2013. p. 717–727.
17. V. Gupta, and G. Lehal, A Survey of Text Summarization Extractive Techniques. Journal of Emerging Technologies in Web Intelligence. 2. 10.4304/jetwi.2.3.258-268, 2010.
18. A. Abuobieda et al. Text summarization features selection method using pseudo genetic-based model. In: Proc. – 2012 int. conf. inf. retr. knowl. manag. CAMP’12. p. 193–197.2012.
19. M.A. Fattah and F. Ren. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Comput Speech Lang 2009;23(1):126–144.
20. K. Nandhini and S.R. Balasundaram. Improving readability through extractive summarization for learners with reading difficulties. Egypt Inform J 2013;14 (3):195–204.
21. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. Comput Networks ISDN Syst 1998;30(1–7):107–117.
22. D. Parveen and H.M. Ramsl and M. Strube. Topical coherence for graph-based extractive summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing. p. 1949–1954.2015.
23. E. Canhasi and I. Kononenko. Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization. Expert Syst Appl 2014;41(2):535–543.
24. G. Salton. et al. Automatic text structuring and summarization. Inf Process Manage;33(2):193–207.1997.
25. O. Medelyan. Computing Lexical Chains with Graph Clustering. In: Proceedings of the ACL 2007 student research workshop. p. 85–90.
26. Y-N Chen, et al. Spoken lecture summarization by random walk over a graph constructed with automatically extracted key terms. In twelfth annual conference of the international speech communication association, 2011.
27. Plaza L, Stevenson M, Díaz A. Resolving ambiguity in biomedical text to improve summarization. Inf Process Manage 2012;48(4):755–66.
28. R. Mihalcea and P. Tarau. A language independent algorithm for single and multiple document summarization. In: Proc. IJCNLP 2005, 2nd int. join conf. nat. lang. process.. p. 19–24.2005.

29. S. Xiong and D. Ji. Query-focused multi-document summarization using hypergraph-based ranking. Inf Process Manage;52(4):670–681.2016.
30. Lin, C.: ROUGE: A Package for Automatic Evaluation of Summaries, In: Text Summarization Branches Out, Association for Computational Linguistics, pp. 74-81.(2004).
31. Wan X., Yang, J.: Improved Affinity Graph Based Multi-Document Summarization, In: Proceedings of the Human Language Technology Conference of the NAACL, Association for Computational Linguistics, pp. 181-184.(2006).
32. K. Sarkar and S. Dam. Exploiting Semantic Term Relations in Text Summarization. International Journal of Information Retrieval Research (IJIRR), 12(1), 1-18. http://doi.org/10.4018/IJIRR.289607.(2022).
33. C. Fellbaum, WordNet: An Electronic Lexical Database. The MIT Press.(1998).
34. T. Mikolov et al. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 3111-3119.(2013).
35. R. Mihalcea, and P. Tarau, "TextRank: Bringing order into texts," In Proceedings of EMNLP2004, 2004.
36. G. Erkan, and D. R. Radev, "LexPageRank: Prestige in MultiDocument Text Summarization," In Proceedings of EMNLP, 2004.
37. X. Wan, and J. Yang, "Improved af¿nity graph based multi-document summarization," In Proceedings of HLT-NAACL, Companion Volume: Short Papers, pp. 181–184, 2006.
38. D. R. Radev, H. Jing, M. Sty, D. Tam, "Centroid-based summarization of multiple documents," Journal of Information Process and Management, 40(6), pp. 919-938, 2004

Authors

Kamal Sarkar K. Sarkar received the B.E degree in Computer Science and Engineering from Jadavpur University, India, in 1996. He has obtained M.E in Computer Science and Engineering from Jadavpur University, India, in 1999. He received his PhD in Computer Science and Engineering from Jadavpur University, India, in 2011. He currently holds Professor with the Department of Computer Science and Engineering, Jadavpur University.

Sohini Roy Chowdhury S. Roychowdhury received the B.Tech degree in computer science and engineering from West Bengal University and Technology, kolkata, West Bengal in 2013, and the M.Tech degree in computer science and engineering from University Of Kalyani,Nadia,West Bengal, in 2016. She is a research fellow of Department of Computer Science and Engineering, Jadavpur University. She currently holds Assistant Professor with the Department of Computer Science and Engineering at RCCIIT, Kolkata. Her research interest includes document retrieval and summarization.