

# Multi-document Text Summarization through Features Relevance Calculation

Verónica Neri-Mendoza, Yulia Ledeneva\*,  
René Arnulfo García-Hernández, Ángel Hernández-Castañeda

Autonomous University of the State of Mexico,  
Toluca, State of Mexico,  
Mexico

veronica.nerimendoza@gmail.com, yledeneva@yahoo.com,  
rearnulfo@hotmail.com, anhernandezc@uaemex.mx

**Abstract.** Multi-document text summarization is obtaining relevant information from a set of documents describing the same topic. However, determining the key sentences in the text to be presented as a summary is difficult. Consequently, it is necessary to use features that help to identify informative sentences from those that are not. However, distinguishing between significant and insignificant features is a challenging task. In this study, we introduced a method to assess the impact of 19 linguistic and statistical features derived from human-written reference summaries. Moreover, we tested them using the DUC01 dataset in two lengths (50 and 100 words). The results demonstrate that the proposed method outperforms state-of-the-art approaches and heuristics based on the ROUGE-1 metric.

**Keywords.** Text features, summarization, multi-document, contribution of features.

## 1 Introduction

Written news is one of the most important forms of expression for citizens to know and understand real-world events. Hundreds of news stories are generated every day, causing information overload. Because of this, it would be easier for readers to read representative fragments of a set of news stories than to read each [29, 20].

Due to various research, it is known that the task of summarization combines important reading and writing skills, as well as the understanding of a large amount of linguistic knowledge [20, 15]. Automatic Text Summarization (ATS) involves extracting the most essential information from a document or a set of documents using advanced methods [2, 24, 23, 7]. There are different classifications of ATS according to how the summary is generated. There are [9, 1, 18]:

**Abstractive:** This approach produces summaries incorporating new content using external resources to interpret the source code.

**Extractive:** In this approach, summaries are produced by weighing sentences to assign a value to each sentence and then selecting the highest values.

**Hybrid:** Summaries are produced by combining the advantages of abstractive and extractive approaches.

Based on the quantity of input documents, summarization techniques can be classified as Single-Document or Multi-Document approaches (ATSMD) [9].

To summarize a text, humans follow the next steps: Read the text, underline the main ideas, and rewrite the main ideas [20, 15]. Commonly, ATS: Commonly, the ATS involves calculating the

relevance of each sentence through text features and selecting the  $k$  sentences with the best relevance as a summary [9, 29].

Research in extractive ATS has explored various features to identify text segments that capture the main idea of a document set. These features are categorized into statistical and linguistic types. Statistical features focus on the distribution of words or topics without interpreting the content of the document, while linguistic features involve applying linguistic knowledge to analyze sentence structures [9, 18].

However, the following questions remain open:

1. What is the contribution of text features?
2. Which sentences will be included?

With this in mind, we examined 19 different statistical and linguistic features and computed the relevance coefficient for each one to assess its contribution. Furthermore, to select the most important sentences, we used a genetic algorithm (GA) to maximize the weight of sentences. Moreover, we tested the proposed method at two different levels of compression: 50 and 100 words. The remainder of the paper is structured as follows: Section 2 reviews the related literature, Section 3 outlines the proposed method, Section 4 presents the experimental results, and Section 5 provides the conclusions.

## 2 Related Works

The effectiveness of features relies on their application and combination to assess the importance of each sentence in the source documents. Assessing the contribution of text features aids in creating a more accurate summary. The two questions mentioned in Section 1 have been addressed as exposed in Sections 2.1 and 2.2.

### 2.1 What is the Contribution of Text Features?

State-of-the-art methods take various approaches to determining the contribution of text features. Some methods evaluate the sentences within input documents and assign a relevance score to each feature.

#### 2.1.1 Scoring from source documents

The importance of features is established from the source documents, with weights assigned to each feature based on the text content. For example, in [14], a straightforward yet competent method for generating summaries through term frequency was introduced. Term frequency generally serves as a criterion for identifying more relevant sentences. The sentences are subsequently ranked based on their scores.

While in [4], the position of sentences and word frequency were initially considered for summarization. Later, in [8], additional text features like key terms and similarity to the title were introduced.

#### 2.1.2 Scoring through coefficient optimization

In [27], sentence extraction was achieved by generating combinations of relevance coefficients. Initially, these coefficients were assigned randomly within the range of 0 to 1, and were subsequently refined using the GA.

In [10], a GA was employed to determine the optimal set of relevance coefficients for ten features, including sentence position, similarity to the title, presence of named entities, and sentence length. The impact of each feature was initially studied to facilitate summary generation. Subsequently, the features were used to train a GA and a mathematical regression algorithm to determine the optimal set of text features and relevance coefficient values.

In the studies analyzed in this section, sentences were evaluated using relevance coefficients, which are integrated into the sentence score by applying the fitness function outlined in Equation 1:

$$F(s) = \sum_{j=1}^C f_j(s), \quad (1)$$

where,  $f_j$  is the relevance coefficient for each  $j$ th feature,  $s$  is the sentence score, and  $C$  is the total number of features.

### 2.1.3 Scoring from manual coefficients

Similar to how relevance is determined based on coefficients computed through optimization, text feature coefficients have also been calculated manually.

In [19], a method was introduced that combines semantic and statistical features, such as key sentences, sentence length, presence of proper nouns, sentence position, similarity to the title, sentence centrality, and inclusion of numbers. During the sentence selection stage, sentences were evaluated using a linear sum, with the coefficients manually determined.

In [16], To evaluate the quality of a summary, an ensemble of features that are both domain—and language-independent was used. These features included similarity to the title, sentence position, sentence length, cohesion between sentences, and coverage. The features were optimized using a memetic algorithm.

In [25], a GA was proposed for generating summaries by selecting sentences using four features: coverage, sentence position, sentence length, and similarity to the title. The results showed enhancements in sentence selection. Nevertheless, the coefficients were manually determined based on the assumption that these values would improve sentence selection. Consequently, these approaches depend on subjective criteria for setting the coefficients.

### 2.2 Which Sentences will be Included?

Generating a summary is a crucial step. The chosen features and their relevance coefficients determine which sentences most effectively describe the document. Various techniques have been applied to this stage in the literature, including decision trees, lexical chains, clustering, latent semantic analysis, neural networks, and optimization methods.

Each of these techniques has its limitations. Clustering is straightforward and intuitive but limits elements to being assigned to one group [5, 9].

Graph-based methods offer understandable models for representing documents but involve complex construction and storage, and they may not accurately capture the definition of words

or sentences [9]. On the other hand, deep learning methods, while effective, need extensive training data [9, 28, 27]. Latent semantic analysis-based methods depend heavily on the grade of the semantic representation of the source documents [5].

Decision trees can only detect sentence associations based on shared phrases [9, 5]. Therefore, it is crucial to determine the contribution of features by deriving weighting coefficients through methods that balance the quality of the summary with the cost of its generation. In ATS research, several datasets, including human-written reference summaries, have been developed to assess the performance of proposed methods. The aim is for the software-produced summaries to be similar to those created by humans.

Despite research into relevance coefficients through optimization or manual assignment, there has been a lack of investigation into using human-written reference summaries as an objective standard for calculating these coefficients in the current state of the art.

## 3 Proposed Method

Given the uncertainty about the usefulness of calculating the contribution of statistical and linguistic text features based on human-written reference summaries, we propose a methodology comprising the following steps: Calculating Text Features, Calculating Relevance Coefficients, concatenating and pre-processing source documents, then performing feature extraction and sentence selection.

### 3.1 Calculating Text Features

The input for this process consisted of human-written reference summaries. These documents were preprocessed through normalization, text derivation, and removal of stopwords. The source documents were tagged with Parts-of-Speech (POS) and Named Entity Recognition (NER) tags. Besides, the content was vectorized using the Word2vec word embedding model to capture word meanings and enhance the linguistic concepts of

the sentences [11]. After this, we consider the following text features:

### 3.1.1 Inclusion of Thematic Words (TW)

TW pertains to topic particular words that frequently appear in the content. In the proposed method, we evaluated rate values ranging from 5% to 15%. Empirically, we observed that using 7% of the most common words. Those that would give a general overview of the documents could be extracted. This feature was calculated using the equation 2:

$$TW(s) = \frac{\text{Number of } TW \in s}{\text{Total Number of } TW}, \quad (2)$$

where the weight of thematic words ( $TW$ ) of sentence  $s$  is equal to the number of thematic words in the sentence  $s$  divided by the total number of thematic words in the document

### 3.1.2 Inclusion of Positive Keywords (PW)

Given that words are the essential components of a sentence, a sentence with more content keywords is considered more important. Therefore, we established positive keywords as the top 7% of the most recurrent words in the documents, as this percentage effectively identifies thematic words. The weight of this feature was calculated using Equation 3:

$$PW(s) = \frac{1}{\text{length}(s)} \sum_{i=1}^n tf_i \cdot P(s \in D|PW), \quad (3)$$

where  $PW(s)$  is the sum of the frequencies of each term considered as a keyword multiplied by its respective probability value in the sentence  $s$ , which belongs to the document  $D$ .

### 3.1.3 Inclusion of Title Words (ITW)

Sentences that contain words from the title may be indicative of the topic of the document and are more likely to be included in the abstract. For this reason, the sentence obtains a high score if it includes words that show in the title of the

document. This feature was calculated using equation 4:

$$ITW(s) = \frac{\text{Words } \in s \cap \text{Words } \in T}{\text{Length}(T)}, \quad (4)$$

where  $ITW(s)$  is obtained by the intersection of the *words* that belong to the sentence  $s$  and the *words* that belong to the title ( $T$ ), between the *length* of the  $T$

### 3.1.4 Inclusion of POS and NER Tagging

The presence of POS or NER tags can indicate the importance of words in a sentence. While it is possible to capture the frequency of all available POS or NER tags (54 in total), we focused on the most common ones (14), which are listed in the following table:

**Table 1.** Description of POS and NER tags

Tag	Description	Tag	Description
CC	Conjunction	DT	Determiner
CD	Cardinal number	JJ	Adjective
VB	Verb base form	IN	Preposition
NN	Singular noun	NNS	Plural noun
VBD	Verb in the past	PER	Personal
VBN	Past participle	DATE	Periods
GPE	Cities and states	ORG	Organizations

The contribution of this feature was calculated using the term frequency of tagged words (see equation 5):

$$TF(t_i) = \frac{t_i}{N}, t_i \in \{POS, NER\}. \quad (5)$$

### 3.1.5 TF-IDF

Term Frequency (TF) estimates how often a word is included in a source document, while Inverse Document Frequency (IDF) considers the number of sentences in which the word appears. A higher TF-IDF value indicates that the word is more frequent in the sentence but less common across the document (see Equation 6):

$$TF - IDF(s) = TF \times (w) \log\left(\frac{N}{L}\right), \quad (6)$$

where  $TF$  = number of times word  $w$  occurs in the document.  $N$  = Total number of documents in the corpus and,  $L$  = Total number of documents in which the word  $w$  occurs.

TF-IDF has the following properties. It assigns a weight to the word  $w$  in document. It is the highest when  $w$  occurs many times within a document and does not occur in the rest of the documents in the corpus. It is lower when  $w$  occurs fewer times in a document or occurs in many documents. It is the lowest when  $w$  occurs virtually in all the documents.

### 3.1.6 Main sentence similarity (SMS)

This feature evaluates the similarity between a sentence  $s$  and the main sentence of the document ( $MS$ ). The use of centrality grows diversity. For this reason, we proposed that the sentence with the highest score in the  $TF - IDF$  calculation be the main sentence,  $MS$ . We computed its similarity to other sentences using cosine similarity and Word2vec vectorization (see equation 7):

$$SMS(s, MS) = \frac{\sum_{i=1}^n s_i SMS_i}{\sqrt{\sum_{i=1}^n s_i} \sqrt{\sum_{i=1}^n MS_i}} \quad (7)$$

### 3.2 Calculating Relevance Coefficients

This step aims to identify the relevance of each feature by computing coefficients based on human-written reference summaries.

Starting from the calculation of the features described above, the following steps were carried out:

1. A feature matrix was created for each human-written reference document. In this matrix, the columns represent the calculated values for each feature ( $C_i$ ), while the rows correlate with the sentences from the source document.
2. The scores obtained by each feature in the document were summed  $\sum_{i=1}^n C_i$ .
3. The average of each feature from human-written reference summaries was computed ( $\frac{\sum(C_i)}{d}$ ), where  $d$  is the number of human-written reference summaries.

4. The relevance coefficients for each feature were calculated from the earlier averages using Bayesian probability. This probability is favorable as it allows the designation of probabilities to individual events and allows the calculation of an event probability based on known probabilities of related events. Equation 8 describes how the relevance coefficients ( $w_i$ ) are computed:

$$relevance(w_i) = \frac{\sum_{i=1}^n C_i}{\sum_{j=1}^m \frac{\sum_{i=1}^n C_i}{d}}, \quad (8)$$

where  $j$  represent the  $j^{th}$  text feature and  $m$  is the number of human reference summaries. As the output of this process, we obtained a vector of relevance coefficients with values ranging from 0 to 1.

### 3.3 Multi-document Summarization Process

This process is initiated with a collection of news documents that need to be summarized (also called source documents). For each collection of source documents, the following processes are applied:

1. **Pre-processing:** The order of the source documents must represent the chronological sequence of events. Therefore, the news in the collection was combined hierarchically to create a meta-document organized from the oldest to the newest news.  
Afterward, the text was normalized through lemmatization, and stopwords were filtered. Moreover, we applied POS and NER tagging. Lastly, the sentences were vectorized using Word2vec to capture word meanings and construct linguistic concepts.
2. **Sentence selection:** The GA was employed to assess and enhance the selection of sentences that will form the summary, like a combinatorial optimization problem.

This process emulates evolution, gradually and repeatedly refining the given target objective. The "strongest" solutions persist, while the "weakest" ones are eliminated [11]:

- **Encoding:** Binary, a gene means a sentence, and the individuals represent candidate summaries.
- **Initial population:** Randomly.
- **Operators:** Selection, crossover, and mutation operators were applied to obtain new solutions.
- **Text Features:** They were computed according to the equations shown in section 3.1. Then, a feature matrix was created. In this matrix, the columns depict to the score values  $S_i$  of features, while the rows correlate with the sentences in the candidate summaries. Since the  $S_i$  scores, a vector of scores was generated employing the equation:  $(\sum S_i)$ .
- **Fitness Function:** The candidate summaries were assessed by equation 9:

$$FA = Max \sum (\sum s_i \times w_i). \quad (9)$$

The fitness of the candidate document  $D_i$  is the maximization of the linear sum of the scores acquired  $\sum s_i$  multiplied by their relevance coefficients  $W = w_i$  (vector obtained in 3.2 section).

- **Stop criteria:** Number of generations.

## 4 Experimental Results

To know the performance of the proposed method. The tests were implemented under the DUC01 dataset. This dataset serves as a point of reference for estimating the quality of summaries and comprises 309 documents ordered in 30 collections. It focuses on English-language news articles. The dataset includes two human-written reference summaries for assessment [1].

Two summaries were created for each collection with compression rates of 50 and 100 words. Moreover, we analyzed the contribution of a group of 19 linguistic and statistical features calculated from the human-written reference summaries.

The contribution percentages of these features in the generated summaries are shown in Fig. 1.

It is observed that the inclusion of thematic words (TW) in the summaries makes a more significant contribution (35%) because when they appear frequently in the documents, they are related to the topic addressed. Consequently, they make the sentence informative.

In contrast, the singular (NN) and plural noun (NNS), personal names (PERSON), organizations (ORG), as well as names of countries, cities, and places (GPE) were features that contributed to the generation of the summaries with, a sum of 21% between them. These features are important because they capture information about who or what about is in the sentence.

One of the important parts of sentences are the determiners (DT), which, together with the verb, personal pronouns or the subject, help give meaning and context to the nouns since they add information about quantity and possession, which is why they represent 4% of relevance in the generation of the summaries.

Adjectives (JJ) that express characteristics or properties attributed to a noun contributed 5%. At the same time, the actions performed by the nouns were captured by the verbs in base form (VB), the verb in the past (VBD), and the past participle (VBN) with a sum of 5% between them.

Regarding the structure of the summaries, the grammatical categories coordinating conjunction had a contribution of 2%, while the prepositions and conjunctions (IN) contributed 5% of relevance. Due to is necessary because they allow the creation of relationships between words and sentences. As for dates (DATE), 2% of relevance was included, while for the cardinal numbers (CD) feature, the contribution was 3% since it can reflect transactions and percentages. The idea is if a sentence contains numerical data, it is important and very likely to be included in the summary.

Finally, Fig. 1 also shows that with regard to TF-IDF, the contribution was 8% relevance. This feature was used to identify the most distinctive thematic features of the documents. In addition, Fig. 1 shows the similarity of the sentence with the main sentence (SMS) contributed 9% of relevance. Finally, the inclusion of title words (ITW) and positive keywords (PW) contributed 2% of relevance each.

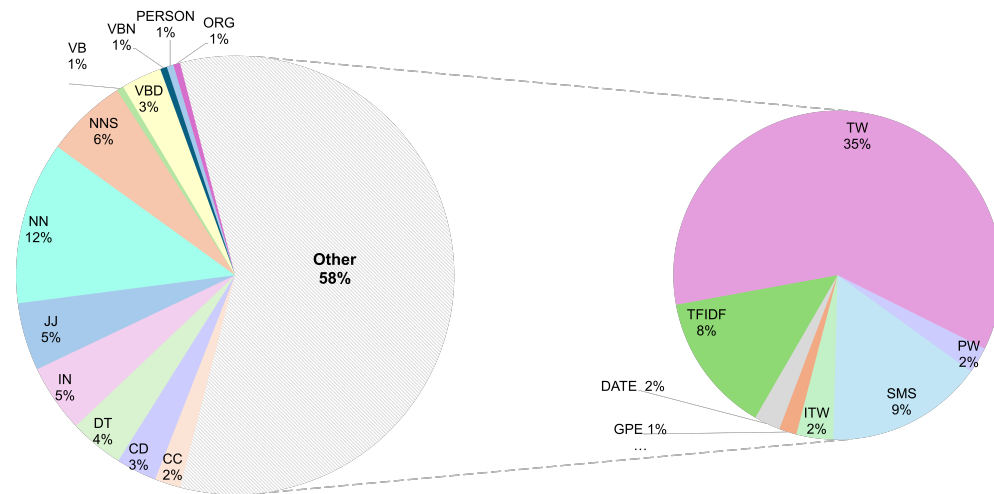


Fig. 1. Percentage distribution of contribution of text features

The GA parameters used to select the sentences that formed the final summaries are shown in the next table:

Table 2. Parameters of GA

Feature	50 words	100 words
Generations	15	85
Population size	$2 \times \text{number of sentences}$	
Elitism	0.03%	
Selection operator	Roulette	
Inversion mutation	0.009%	

As observed in Table 2, the number of generations varied according to the summary length. The longer the length, the more generations were required. Moreover, the best results were obtained with the selection operator roulette operator.

We utilized the ROUGE system to assess the summaries produced by the proposed method. This system measures the quality of the generated summaries by comparing them with human-written reference summaries using n-grams. Specifically, we emphasized using ROUGE-1 and ROUGE-2,

which are widely regarded as a dependable metric for this type of evaluation [7].

The heuristics used to contrast the performance of the proposed method are outlined below.

**Topline:** Consists of obtaining the best selection of sentences (via GA) according to their similarity concerning human-written reference (ideal) summaries. Therefore, these summaries are a reference point that any ATSMMD method aspires to achieve, even if there is disagreement among ideal summaries [21].

**BF:** The Baseline-first (BF) selects the first sentences from source documents sorted chronologically, generating extractive summaries according to the number of words.

**BR:** The Baseline-random (BR) randomly selects sentences from source documents till the required length is complete to include them as a summary.

**BFD:** Baseline-first-document (BFD) takes out the first sentences from the earliest document until the required summary length is reached.

**LB:** Lead Baseline (LB) incorporates the first 50 and 100 words of the most recent document as a summary. Likewise, the input documents must be chronologically sorted.

**Table 3.** Comparison with heuristics and state-of-the-art methods 50 Words

Method	ROUGE-1	Advance (%)	ROUGE-2	Advance (%)
Topline	40.395 (1)	100.000 %	15.648(1)	100.000 %
GA	28.023 (2)	39.258%	6.272 (2)	31.656 %
Proposed	27.854 (3)	38.427%	4.699 (3)	20.190 %
RBM	27.369 (4)	36.046%	4.617 (4)	19.593%
BFD	25.435 (5)	26.551%	4.301 (7)	17.289%
BF	25.194 (6)	25.368%	4.596 (5)	19.440%
Baldwin	22.906 (7)	14.134%	3.054 (8)	8.200%
CBA	22.679 (8)	13.020%	2.859 (10)	6.778%
LB	22.620 (9)	12.730%	4.341 (6)	17.581%
NeATS	22.594 (10)	12.603%	2.963 (9)	7.536%
BR	20.027 (11)	0.000%	1.929 (11)	0.000%

**Table 4.** Comparison with heuristics and state-of-the-art methods 100 Words

Method	ROUGE-1	Advance (%)	ROUGE-2	Advance (%)
Topline	47.256 (1)	100.000 %	18.994(1)	100.000 %
Proposed	34.053 (1)	34.838 %	7.632 (2)	27.708%
GA	33.985 (2)	34.503 %	7.617 (3)	27.613%
RBM	32.923 (3)	29.261 %	6.985 (4)	23.592%
BF	31.716 (4)	23.304 %	6.962 (5)	23.445%
BFD	30.462 (5)	17.115 %	5.962 (6)	17.083%
Baldwin	28.647 (6)	8.158 %	4.760 (7)	9.435%
NeATS	28.195 (7)	5.927 %	4.037 (9)	4.835%
LB	28.195 (8)	5.927 %	4.109 (8)	5.293%
BR	26.994 (9)	0.000 %	3.277 (11)	0.000%
CBA	26.741 (10)	-1.248%	3.510 (10)	1.482%

Subsequently, we present the state-of-the-art techniques used to compare the performance of the proposed method.

**CBA:** The Clustering-Based Approach (CBA) creates summaries using sentences as topics. The topics are then clustered using two types of clustering: hierarchical and partitioning (K-means). Finally, the most relevant topics are selected for the final summary [6].

**NeATS:** NeATS is a method that employs term clustering (also known as the “buddy system”) to match sentences to select the most relevant sentences from source documents [12].

**GA:** The authors in [17] proposed a GA to optimize sentence selection using Coverage and Sentence position.

**RBM:** This method proposes using the Restricted Boltzmann Machine (RBM) to identify the relationships among nine text features. These features include TF-IDF, SMS, POS, NER, and Sentence Length [26].

**Baldwin:** This method employs sentence selection using entropy [3]. Therefore, a sentence concerning the collection of documents is relevant if it contains words of low entropy.

Tables 3 and 4 compare the proposed method with state-of-the-art techniques and heuristics.



**Table 5.** Resulting ranking of the methods

Method	Position											Result Rank
	1	2	3	4	5	6	7	8	9	10	11	
Topline	4	0	0	0	0	0	0	0	0	0	0	4.000 (1)
Proposed	2	0	2	0	0	0	0	0	0	0	0	3.636 (2)
GA	2	0	2	0	0	0	0	0	0	0	0	3.636 (2)
RBM	0	0	0	4	0	0	0	0	0	0	0	2.909 (3)
BFD	0	0	0	0	1	2	1	0	0	0	0	2.181 (5)
BF	0	0	0	0	3	1	0	0	0	0	0	2.454 (4)
Baldwin	0	0	0	0	0	0	3	1	0	0	0	1.727 (6)
CBA	0	0	0	0	0	0	0	1	0	2	1	0.818 (9)
LB	0	0	0	0	0	1	0	1	2	0	0	1.454 (7)
NeATS	0	0	0	0	0	0	0	1	2	1	0	1.090 (8)
BR	0	0	0	0	0	0	0	0	0	1	3	0.454 (10)

Additionally, we calculated the improvement in the summarization task, considering that any method cannot perform worse than randomly selecting sentences (BR), which is set to 0%.

The best possible performance, referred to as the Topline, is set at 100%. By utilizing BR and the Topline, we can recalculate the F-measure results to assess the improvement relative to the worst and best results. This advancement is displayed in the third and fifth columns of the tables. The number in parentheses within each table slot indicates the ranking of each method.

The ROUGE-1 and ROUGE-2 scores demonstrate that the proposed method surpasses all state-of-the-art methods and heuristics for the lengths of summaries of 100 words. Although the proposed method does not outperform the GA method in 50 words summaries, it achieves a comparable value, indicating a promising gap for future research. Additionally, the method enhances sentence selection overall. Furthermore, the proposed method shows close performance to the Topline, highlighting the extent of the achieved improvement.

To consolidate all the results from ROUGE-1 and ROUGE-2 for 50 and 100 words, Table 5 presents them in a unified format, ranking them based on

Equation 10, which has been applied in [17]:

$$Rank (method) = \sum_{r=1}^{11} \frac{(11 - r + 1)R}{11}, \quad (10)$$

where  $R$  indicates the number of times the method appears in the  $r$ -th rank.

Table 5 provides a comprehensive comparison of the summarization methods. From the findings, we can note that the BR demonstrates the lowest performance.

Meanwhile, both the proposed method and the GA significantly enhance results, achieving second place in the ranking.

## 5 Conclusions

In existing literature, human-written reference summaries have typically been used to evaluate the performance of proposed methods, not to determine the score of the features.

Our findings indicate that thematic words are the most influential feature in summary generation, with features related to nouns, verbs, and adjectives also playing significant roles. Additionally, we evaluated the contribution of features related to grammatical categories, such as determiners, conjunctions, and prepositions.

After determining the contribution of each feature, we optimized sentence selection using GA. The results demonstrate an enhancement in sentence selection across various summary lengths, as indicated by the ROUGE-1 and ROUGE-2 measures.

The contribution derived from human-written reference summaries offers a valuable starting point for assigning relevance to features, offering practical insights for future research and development. However, since ROUGE depends on human-written summaries for evaluation, it is crucial to assess the performance of methods using evaluation techniques that do not rely on human references [13, 22].

## References

1. **Akhmetov, I., Nurlybayeva, S., Ualiyeva, I., Pak, A., Gelbukh, A., Akhmetov, I., Nurlybayeva, S., Ualiyeva, I., Pak, A., Gelbukh, A. (2023).** A comprehensive review on automatic text summarization. *Computación y Sistemas*, Vol. 27, No. 4, pp. 1203–1240.
2. **AL-Khassawneh, Y. A., Hanandeh, E. S. (2023).** Extractive Arabic text summarization-graph-based approach. *Electronics* 2023, Vol. 12, Page 437, Vol. 12, No. 2, pp. 437. DOI: 10.3390/ELECTRONICS12020437.
3. **Baldwin, B., Ross, A. (2001).** Baldwin language technology's DUC summarization system. *Proceedings of the 1st Document Understanding Conference*, New Orleans, LA.
4. **Baxendale, P. B. (1958).** Machine-Made Index for Technical Literature—An Experiment. *IBM Journal of Research and Development*, Vol. 2, No. 4, pp. 354–361. DOI: 10.1147/rd.24.0354.
5. **Belwal, R. C., Rai, S., Gupta, A. (2023).** Extractive text summarization using clustering-based topic modeling. *Soft Computing*, Vol. 27, No. 7, pp. 3965–3982. DOI: 10.1007/S00500-022-07534-6/METRICS.
6. **Boros, E., Kantor, P. B., Neu, D. J. (2001).** A Clustering Based Approach to Creating Multi-Document Summaries. Technical report.
7. **Dhaini, M., Erdogan, E., Bakshi, S., Kasneci, G. (2024).** Explainability meets text summarization: A survey. *International Natural Language Generation Conference*, pp. 631–645.
8. **Edmundson, H. P. (1969).** New Methods in Automatic Extracting. *Journal of the ACM*, Vol. 16, No. 2, pp. 264–285. DOI: 10.1145/321510.321519.
9. **El-Kassas, W. S., Salama, C. R., Rafea, A. A., Mohamed, H. K. (2021).** Automatic text summarization: A comprehensive survey. DOI: 10.1016/j.eswa.2020.113679.
10. **Fattah, M. A., Fattah, M. A. (2014).** A hybrid machine learning model for multi-document summarization. *Appl Intell*, Vol. 40, pp. 592–600. DOI: 10.1007/s10489-013-0490-0.
11. **Kostiuk, Y., Pichardo-Lagunas, O., Mandii, A., Sidorov, G. (2023).** Automatic detection of semantic primitives using optimization based on genetic algorithm. *PeerJ Computer Science*, Vol. 9, pp. e1282. DOI: 10.7717/PEERJ-CS.1282.
12. **Lin, C.-Y., Hovy, E., .** NeATS in DUC 2002. Technical report.
13. **Louis, A., Nenkova, A. (2013).** Automatically Assessing Machine Summary Content Without a Gold Standard. *Computational Linguistics*, Vol. 39, No. 2, pp. 267–300. DOI: 10.1162/COLI\_A\_00123.
14. **Luhn, H. P. (1958).** The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159–165. DOI: 10.1147/rd.22.0159.
15. **Mahlow, C., Ulasik, M. A., Tuggener, D. (2024).** Extraction of transforming sequences and sentence histories from writing process data: a first step towards linguistic modeling of writing. *Reading and Writing*, Vol. 37, No. 2, pp. 443–482. DOI: 10.1007/s11145-021-10234-6.
16. **Mendoza, M., Cobos, C., León, E., Lozano, M., Rodríguez, F., Herrera-Viedma, E. (2014).** A New Memetic Algorithm

- for Multi-document Summarization Based on CHC Algorithm and Greedy Search. Springer, Cham, pp. 125–138. DOI: 10.1007/978-3-319-13647-9\_14.
17. **Neri-Mendoza, V., Ledeneva, Y., García-Hernández, R. A. (2020).** Unsupervised extractive multi-document text summarization using a genetic algorithm. *Journal of Intelligent & Fuzzy Systems*, Vol. 39, No. 2, pp. 2397–2408. DOI: 10.3233/JIFS-179900.
  18. **Neri Mendoza, V., Ledeneva, Y., García-Hernández, R. A., Hernández Castañeda, Á. (2024).** Relevance of sentence features for multi-document text summarization using human-written reference summaries. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 14755 LNCS, pp. 319–330. DOI: 10.1007/texts1ash978-3-031-62836-8\_30.
  19. **Qaroush, A., Abu Farha, I., Ghanem, W., Washaha, M., Maali, E. (2021).** An efficient single document Arabic text summarization using a combination of statistical and semantic features. *Journal of King Saud University - Computer and Information Sciences*, Vol. 33, No. 6, pp. 677–692. DOI: 10.1016/J.JKSUCI.2019.03.010.
  20. **Rezaei, H., Amid Moeinzadeh Mirhosseini, S., Shahgholian, A., Saraee, M., Shahgholian AShahgholian, A., Rezaei, H. (123).** Features in extractive supervised single-document summarization: case of Persian news. *Language Resources and Evaluation*. DOI: 10.1007/s10579-024-09739-7.
  21. **Rojas-Simón, J., Ledeneva, Y., García-Hernández, R. A. (2018).** Calculating the significance of automatic extractive text summarization using a genetic algorithm. *Journal of Intelligent and Fuzzy Systems*, Vol. 35, No. 1, pp. 293–304. DOI: 10.3233/JIFS-169588.
  22. **Rojas-Simón, J., Ledeneva, Y., García-Hernández, R. A. (2021).** Evaluation of text summaries without human references based on the linear optimization of content metrics using a genetic algorithm. *Expert Systems with Applications*, Vol. 167, pp. 113827. DOI: 10.1016/J.ESWA.2020.113827.
  23. **Supriyono, Wibawa, A. P., Suyono, Kurniawan, F. (2024).** A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal*, Vol. 7, pp. 100070. DOI: 10.1016/J.NLP.2024.100070.
  24. **Torres-Moreno Juan-Manuel (2014).** *Cognitive Science and Knowledge Management Series. Automatic Text Summarization.*
  25. **Vázquez, E., García-Hernández, R. A., Ledeneva, Y. (2018).** Sentence features relevance for extractive text summarization using genetic algorithms. *Journal of Intelligent & Fuzzy Systems*, Vol. 35, No. 1, pp. 353–365. DOI: 10.3233/JIFS-169594.
  26. **Verma, P., Om, H. (2019).** MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization. *Expert Systems with Applications*, Vol. 120. DOI: 10.1016/j.eswa.2018.11.022.
  27. **Verma, S., Vagisha Nidhi (2019).** *Extractive Text Summarization Using Deep Learning. Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, pp. 43–56. DOI: 10.1109/ICCUBEA.2018.8697465.
  28. **Xiong, Y., Yan, M., Hu, X., Ren, C., Tian, H. (2023).** An unsupervised opinion summarization model fused joint attention and dictionary learning. *Journal of Supercomputing*, Vol. 79, No. 16, pp. 17759–17783. DOI: 10.1007/S11227-023-05316-X/METRICS.
  29. **Yang, Y., Tan, Y., Min, J., Huang, Z. (2024).** Automatic text summarization for government news reports based on multiple features. *Journal of Supercomputing*, Vol. 80, No. 3, pp. 3212–3228. DOI: 10.1007/S11227-023-05599-0/METRICS.

Article received on 23/03/2024; accepted on 02/06/2024.

\*Corresponding author is Yulia Ledeneva.