Multimedia Tools and Applications (2024) 83:23737–23759
https://doi.org/10.1007/s11042-023-16358-x

A binary grey wolf optimizer to solve the scientific document summarization problem

Ranjita Das[1,2] · Dipanwita Debnath[2,3] · Partha Pakray[4] · Naga Chaitanya Kumar[2]

Received: 1 February 2022 / Revised: 28 June 2023 / Accepted: 17 July 2023 /
Published online: 17 August 2023

Abstract

The extraction of information from the extensive volume of online textual data poses a significant challenge, and text summarization plays a pivotal role in overcoming this challenge. Conventionally, an extractive text summary, which consists of the most relevant sentences from the text itself, can effectively represent the given text. However, identifying such a subset of sentences is challenging. To overcome this problem, this paper introduces a Binary Gray Wolf Optimization (BGWO)-based text summarization approach that tackles the sentence selection problem. The proposed system performs pre-processing on the input texts, identifies noteworthy features, and generates a text segment that includes the most relevant sentences. Subsequently, the BGWO algorithm is employed to generate an optimal summary from the text segment. The BGWO-based text summarization approach begins by initializing the population as a set of feasible solution vectors represented by binary values, indicating the presence or absence of sentences in the summary. Thereafter, fitness functions incorporating textual features are constructed, and the population's fitness is evaluated. Through non-dominated sorting and crowding distance, individuals are categorized as alpha, beta, delta, or gamma wolves. In each iteration, their positions are updated using crossover and mutation operations, and individuals are ranked based on their fitness scores. Finally, the alpha wolf is selected as the optimal summary candidate. The proposed method is evaluated and compared to various existing methods on the DUC-2001, DUC-2002, and ScisummNet datasets using ROUGE measures. Statistics based on ROUGE scores are also computed, demonstrating that the proposed system is statistically significant. The experimental results demonstrate that BGWO outperforms ROUGE scores for single-document summarization, including low-resource documents.

Keywords ScisummNet · Grey wolf optimization · Extractive text summarization · News summarization · Scientific document summarization · Research article summarization

Dipanwita Debnath, Partha Pakray and Naga Chaitanya Kumar are contributed equally to this work.

✉ Ranjita Das
ranjita.nitm@gmail.com

Extended author information available on the last page of the article

1 Introduction

The incredible growth of data on the Internet has made it difficult to obtain information from them in the allotted period of time [2]. Examples include extracting significant information from news articles and scientific research papers. While both scientific and news documents are single-text documents, news reports are typically concise, consisting of a few paragraphs with a generic nature. On the other hand, scientific documents are well-structured, encompassing sections such as abstracts, introductions, and results, often containing lengthy paragraphs. Scientific documents also exhibit more redundancy and citations to other researchers' work. Due to limited resources, the development of a supervised abstractive summarization system for scientific documents, which necessitates a large training dataset, remains challenging [22, 34]. However, sentence-based extractive text summarization systems, specifically optimization-based unsupervised approaches, have shown promising performance in addressing this challenge. These systems have evolved and gained popularity in recent years due to their ability to effectively tackle such problems. They are characterized by the application of rules derived from user-specific requirements [17, 20, 21, 30].

Optimization-based approaches in Text Summarization (TS) aim to generate high-quality summaries by selecting the most relevant sentences and presenting them effectively. Several techniques have been proposed to enhance the performance of these approaches. Ghodratnama et al. [15] employ feature engineering to identify weighted features for the objective function formulation, combined with k-means clustering and K-nearest neighbor (K-NN) classification algorithms. To generate summaries, a semantic graph-based representation of text (KUSH) and the Maximum Independent Set are used. A decomposition-based multi-objective artificial bee colony (MOABC/D) is proposed to generate summaries automatically by maximizing two objective functions: content coverage and redundancy [33]. D.debnath et al. [10] used an archive-based micro genetic algorithm to generate extractive summaries by formulating the objective functions considering informative, content coverage, and redundancy aspects. In [11], also, objective functions are formed considering informative, content coverage, and redundancy aspects. However, a modified cat swarm optimization approach is used to solve the summarization problem. Genetic Algorithm (GA) is used as an underlying strategy for multi-objective TS designed by E. Vazquez et al. [36], where a single objective function considering the weighted sum of different sentence feature scores is optimized. Aliguliyev et al. proposed COSUM [2], where an adaptive binary Differential Evolution (DE)-based optimization method is used to detect the number of optimal clusters. After that, the K-mean clustering technique is used for summary generation. N. Saini et al. in ESDS-SMODE [32] used a self-organizing map incorporating the DE approach to solve the single document TS problem; two benchmark clustering indexes were optimized to detect the optimal number of clusters. Sentences are then selected from the clusters, considering the weighted sentence scores. In [31], six features are used as objective functions, and a differential evolutionary-based approach is used. The majority of the above mentioned optimization approaches are for short documents [2, 10, 31, 32], having certain drawbacks: (i) The use of cosine similarity measures [2, 5, 33] as a similarity metric, may have limitations in accurately capturing the semantic similarity between sentences. (ii) Poor ROUGE scores [2, 10, 31, 32], indicating sub-optimal performance in terms of summarization quality and content coverage, (iii) The utilization of optimization techniques to detect the number of clusters [2, 32], can introduce additional complexity and potential challenges in determining the appropriate number of clusters for effective summarization.

In the domain of Scientific Document Summarization (SDS), various works have focused on aspects such as citation community identification, citation type classification, citation quality assessment, and the motives behind citations [14, 16, 37]. The effectiveness of citations in scientific papers has been investigated in [6], and datasets from the CL-SciSumm workshop (2016-2020) have been released to promote research in citation contextualization [18, 19]. Here, we discuss some notable approaches in this field.

In [5], citation-related sentences are extracted using cosine similarity, numerically represented based on their similarity scores, and ranked using the Support Vector Machine (SVM) approach. High-ranked sentences are then used for summary generation. Similarly, [23] employs cosine similarity to extract citation-related text spans, followed by lexicon-based sentence representation and ranking using an SVM classifier for summary formation. In [34], cosine similarity is used to generate the text span of the reference paper, and a recurrent neural network model is employed for summary generation.

The BUPT system [40] combines Word2Vec and pre-trained encoders (BERT) with various neural networks to create citation-based text spans, which are then used for summary generation using a neural network-based model. SERGE [12] utilizes a BERT-based ensemble model to select the most salient sentence, and correlation analysis is performed to establish similarity between the citing snippet and the title/abstract. Several neural network-based models such as BART [22, 29], Pegasus, and Longformer [4] have been designed for summary generation from text documents. NUDT [38] combines features-based random forest, BM25, VSM, and a voting strategy for summary generation. A BERT architecture-based scientific document summarization approach is proposed in [29] for aiding the COVID-19 crisis. The model continuously learns from online data to minimize catastrophic forgetting. In NJUST [26], a weighted voting-based linear SVM is employed to identify the reference span, while facet-based clustering is applied for summary generation. NLP-NITMZ [8] extracts citation-related text spans using cosine and JACCARD similarity measures, represents the text span as a feature vector, and applies rule-based sentence selection techniques for summary generation. Similarly, [9] uses a DE-based optimization approach for summary generation with a similar text span representation. However, many of these approaches suffer from limitations. Firstly, several are supervised systems trained on low-resource datasets, resulting in improper and unreadable summaries [8]. Over-fitting problems are also observed in some systems. Additionally, pre-processing steps such as removing sentences related to citations are not performed consistently. Furthermore, most optimization-based SDS systems rely on syntactic similarity measures, such as cosine similarity, for extracting similar sentences.

Motivated by the fact that the SDS holds immense potential for decision making and knowledge acquisition, an optimization-based TS system provides the most promising results, and the Grey Wolf Optimizer is one of them. In this paper, one such binary variant of GWO [13, 35], is used. In this study, a specific binary variant of GWO, referred to as Binary Grey Wolf Optimizer (BGWO) [13, 35], is employed. The choice of BGWO is motivated by its computational efficiency, as it generates offspring solely based on the alpha, beta, and delta wolves. Moreover, BGWO incorporates a leader enhancement strategy to track the global optimum and prevent the algorithm from getting trapped in local optima.

In this paper, we propose an extractive Text Summarization (TS) approach for single-document, single-lingual (English) texts. We evaluate our approach using the DUC-2001, DUC-2002, and ScisummNet datasets. DUC-2001 and DUC-2002 are used directly to generate generic summaries after pre-processing, as they contain text and corresponding summary pairs. For ScisummNet, which consists of scientific documents, we create summaries based on citations. Each folder in ScisummNet contains a reference paper (RP), citations (sentences) from RP's citing papers (CP), and corresponding summaries. So the following steps are performed to generate summaries from these documents:

1. Text-span creation: In the domain of Scientific Document Summarization (SDS), the process of text-span creation involves evaluating the similarity score between each sentence of the reference paper (RP) and all its corresponding citing sentences. If a sentence is found to have a similarity score of 80% or higher with one or more citations, it is included in the text-span. This step is crucial in reducing the amount of text that needs to be considered for generating the summary. However, for datasets like DUC-2001 and DUC-2002, the original text itself is used as the text-span from which the summary is generated. Similarly, in cases where an insufficient number of citations is available, the entire scientific document can be treated as the text-span for summary generation.
2. Pre-processing: During the pre-processing stage, the text is extracted from the documents and divided into individual sentences, which are then tokenized. Unwanted elements and noises are eliminated, and the sentences are converted to lowercase. In the case of scientific documents (reference papers), regular expressions are utilized to identify and exclude sentences that include citations to the work of other researchers, as they are not relevant for the summarization process.
3. Formulation of fitness functions: Text summarization pursues multiple objectives, including content coverage, informativeness, anti-redundancy, and citation similarity in citation-based SDS. To tackle these aspects, fitness or objective functions are developed. These functions result in multiple objective or fitness functions, allowing for a more accurate evaluation of summary quality. By integrating these functions, the summarization system can optimize the summary generation process and effectively accomplish the intended summarization goals.
4. Summary generation using GWO approach: The grey wolf optimization is used to generate summaries whose internal steps are discussed in Section 3.

Experiments are carried out and tested on DUC-2001, DUC-2002, and ScisummNet data sets by varying the objective/fitness functions. Recall-Oriented Understudy for Gisting Evaluation [24] (ROUGE) measures are used to evaluate the outputs. ROUGE includes measures for determining the quality of system-generated summaries by comparing them to other target summaries. This paper's main contributions are as follows:

1. A novel approach for scientific document summarization is proposed, which is based on citation-oriented summarization using the Grey Wolf Optimization (GWO) algorithm. This approach is applicable for a wide range of summarization purposes, including generic, query-based, low or high resource, and short or long text summarization. Its versatility allows for various applications in different domains.
2. Pre-processing steps are performed, including the removal of sentences that contain citations to other researchers' work. This ensures that the summarization process focuses on the content of the document itself and excludes references to external sources.
3. The proposed approach addresses the multi-objective nature of scientific documents by formulating and presenting multiple fitness functions. These fitness functions enable the optimization of summary generation considering various objectives such as content coverage, informativeness, anti-redundancy, and citation similarity.
4. Experiments on fitness function design are conducted to demonstrate their effectiveness in achieving high-quality multi-objective summarization.

The paper is organized as follows: Section 2 provides an in-depth discussion on feature extraction and representation, as well as Earth Mover Distance, Non-dominated sorting &

crowding distance of NSGA-II, and an overview of the binary grey wolf optimization algorithm. Section 3 presents the detailed system architecture. In Section 4, the corpora and experimental setups are briefly explained, along with a comprehensive analysis of the system results. Lastly, Section 5 concludes the paper by presenting some ideas for future research.

2 Methodology and feature extraction

This section includes information about the algorithm used, the employed EMD similarity measure, the non-dominated sorting & crowding distance of NSGA-II, and the details of the features utilized in the study.

2.1 Earth Mover Distance (EMD)

The EMD [39] between sentences are calculated based on the many-to-many semantic distances between their words. For this purpose a lexical database WordNet [27] is used. We have used the EMD measure to calculate the features scores, such as sentence-similarity-with-the-title-of-the-document, anti-redundancy score, and cohesion.

EMD stands for Earth Mover's Distance, also known as Wasserstein distance. It is a measure of dissimilarity between two probability distributions. The EMD quantifies the minimum amount of work required to transform one distribution into another, where the work refers to the effort needed to move mass from one location to another. Mathematically, the Earth Mover's Distance between two probability distributions P and Q can be expressed as shown in the (1):

$$\mathrm{EMD}(P, Q) = \min \sum_{i=1}^{n} \sum_{j=1}^{m} d(i, j) \cdot f(i, j) \tag{1}$$

where i and j represent the locations or bins in the distributions P and Q, d(i, j) is the distance between locations i and j, and f(i, j) represents the amount of mass moved from location i to location j. The summation is taken over all pairs of locations in the distributions.

The EMD considers both the distances between locations and the amount of mass that needs to be transported, providing a more nuanced measure of dissimilarity between distributions compared to other metrics like Euclidean distance or Kullback-Leibler divergence. It has found applications in various fields, including computer vision, image matching, and natural language processing.

2.2 Non-dominated sorting & crowding distance

In order to handle the conflicting objectives, NSGA-II's [7] non-dominated sorting and crowding distance are employed for sorting the Pareto optimal fonts, which efficiently explores and maintains a diverse set of non-dominated solutions.

Non-dominated sorting categorizes individuals into different fronts based on their dominance relationship. An individual is considered non-dominated if there is no other individual that outperforms it in all objectives. NSGA-II assigns a rank to each individual based on their dominance, where individuals in higher ranks are more preferable.

Crowding distance is a diversity-preserving mechanism employed in NSGA-II. It quantifies the degree of crowding around an individual within a front. It measures the density of individuals in the objective space, aiming to maintain a well-distributed set of solutions. By considering the crowding distance, NSGA-II can balance exploration and exploitation, promoting diversity in the final Pareto front.

2.3 Features extraction and representation

The identification, selection, and representation of relevant features is critical in the TS system. In this system, features such as, a number-of-informative-words (F_1), uni-gram based Term- Frequency and Inverse-Document-Frequency (TF-IDF) score (F_2), bi-gram based TF-IDF score (F_3), sentence-similarity-with-the-title-of-the-document (F_4), anti-redundancy(F_5) and cohesion (F_6) are utilized as the prime features. However, based on the data set orientation, two additional features are considered for scientific documents. The ScisummNet dataset includes reference papers (RP), a list of citation sentences from RP's citing papers (CP), and citation-based summaries for each reference paper. Reference papers are those scientific papers whose summaries has to be created, and citing papers are those scientific papers, which has cited RP in their paper. Concerning citation-baed SDS, we identified two relevant features (F_7), and (F_8), namely, citation-similarity score and number-of-citation score. All of these features are described in detail below.

1. Number-of-informative-words (F_1): The Number-of-informative-words feature (F_1) represents the count of words in a sentence after preprocessing, which involves the removal of stop words and other noise elements from summary sentences, as described in [10]. While other feature scores (F_2-F_6) are calculated within the range of [0,1], we apply feature scaling, specifically the min-max normalization technique, to represent F_1 within the same range. The new F_1 score for a sentence, denoted as F_{1new}, is computed using (2), where F_{1Min} and F_{1Max} denote the lower and upper bounds of the range, respectively.

$$F_{1new} = \frac{F_1 - F_{1Min}}{F_{1Max} - F_{1Min}} \tag{2}$$

2. n-gram based TF-IDF score (F_2, F_3): the Term Frequency-Inverse Document Frequency (TF-IDF) score of a sentence is a numerical representation that measures the importance of a specific n-gram (word or phrase) within a sentence, considering its frequency in the sentence and across the entire document collection. Higher TF-IDF scores indicate greater relevance and importance of the n-gram in the context of the sentence and the document collection. We calculated the TF-IDF for each sentence up to 2 grams because 2 grams also aids in the extraction of more relevant sentences.
   The TF component calculates the relative frequency of an n-gram within a sentence. It is computed by dividing the number of times the n-gram appears in the sentence by the total number of n-grams in the sentence, as shown in the (3):

$$TF_{n-gram} = \frac{\text{Number of occurrences of the n-gram in the sentence}}{\text{Total number of n-grams in the sentence}} \tag{3}$$

   The IDF component measures the inverse document frequency of an n-gram across the entire document collection. It is calculated as the logarithm of the ratio between the total number of documents in the collection and the number of documents that contain the n-gram, as shown in the (4):

$$IDF_{n-gram} = log \frac{\text{Total number of documents}}{\text{Number of documents containing the n-gram}} \tag{4}$$

The final TF-IDF score for the n-gram in the sentence is obtained by multiplying the TF and IDF values as shown in the (5), and all the n-gram TF-IDF scores are summed up to get a single score per sentence.

$$TF - IDF_{n-gram} = TF_{n-gram} * IDF_{n-gram} \tag{5}$$

3. Sentence-similarity-with-the-title-of-the-document (F_4): These feature (F_4) quantifies the extent to which a summary sentence covers the content of the document's title or headline. It is calculated by measuring the similarity between each summary sentence (S_i) and the document's title (T) using (6).

$$F_4 = \sum_{i=1}^{N} EMD_{sim}(S_i, T) \tag{6}$$

4. Anti-redundancy (F_5) is a feature that quantifies the presence of redundant sentences in a summary. A lower F_5 score indicates a higher number of redundant sentences, while a higher score signifies a lower level of redundancy. To compute F_5, we use (7), where S_i and S_j represent two summary sentences, N represents the total number of sentences in the summary , and a threshold of 0.25 (i.e. 75% similarity as similar). This feature is designed as a maximization criteria.

$$F_5 = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \begin{cases} 0, & \text{if } EMD_{sim}(S_i, S_j) leq 0.25 \\ 1, & \text{otherwise} \end{cases}}{N} \tag{7}$$

5. Cohesion (F_6) is a measure of readability, ensuring that each summary sentence is connected to its following sentence. We evaluate cohesion using (8), where F_6 is computed as the sum of indicator functions that compare the Earth Mover's Distance similarity (EMD_{sim}) between adjacent sentences S_i and S_j, with a threshold of 0.50.

$$F_6 = \sum_{i=1;\, j=i+1}^{N} \begin{cases} 1, & \text{if } EMD_{sim}(S_i, S_j) \leq 0.50 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

6. Citation-similarity score(F_7): From the citation papers related to a reference paper, citing texts are extracted. For each reference paper's sentences, their similarity score with respect to all the citations are calculated and the best similarity score is used as the citation-similarity score(F_7).
7. Number-of-citation score (F_8): The Number-of-citation score (F_8) is computed for each sentence in the reference papers. It represents the total number of citations received concerning a sentence and is calculated using the EMD similarity measure. If a sentence is found to be 75% or more similar to a citation, they are regarded as similar, and F_8 is determined as the total count of such similarities.

To measure similarity or dissimilarity, we employ the Earth Mover Distance (EMD) similarity measure [3]. It is worth noting that all these features, including F_8, are designed as maximization criteria, where higher feature scores indicate higher fitness scores for the individuals.

2.4 Binary grey wolf optimizer

The Grey Wolf Optimizer (GWO) is a swarm-based meta-heuristic optimization method developed by Mirjalili et al. in 2014 [28], inspired by the hunting behavior of grey wolves. In GWO, the alpha (α) wolf, representing the fittest individual, serves as the prominent decision-making leader. The beta (β) and delta (δ) wolves, ranking second and third in fitness, respectively, act as assistant decision-makers, while the remaining wolves are referred to as omega (ω) wolves [35].

In our approach, we incorporate GWO into the text summarization problem by utilizing crossover and mutation operations in the position updating strategy. The wolf population's positions are iteratively updated, followed by fitness reassessment and re-ranking. Ultimately, the alpha (α) wolf is considered the optimal candidate for the summarization task.

3 Proposed system

This section discusses the proposed extractive single-document summarization approach utilizing grey wolf optimizer (ESDS-BGWO), an automatic summarization system. The pseudo-code of the proposed system is shown in Algorithm 1 and flowchart is presented in Fig. 1. Each conceptual step, including pre-processing of data, are enumerated below.

1. Pre-processing: pre-processing is a crucial step in our approach, as it involves generating a text-span from the reference paper using its citing papers citations. The text-span comprises relevant sentences related to the citing sentences in the paper. To accomplish this, we calculate the similarity scores between each sentence in the reference paper and its corresponding citing sentences. If the similarity score exceeds 75%, the sentence is added to the text-span. The pre-processing of the dataset involves several steps, as illustrated in Fig. 2.

   (a) Extraction and segmentation: The reference paper, and citations are extracted and segmented into individual sentences. Tokenization is performed to break down the sentences into individual tokens.
   (b) Removal of citations: In scientific documents (reference papers), sentences containing citations to other works are eliminated using regular expressions. This helps in excluding sentences that refer to external sources, such as "[9]," "XYX et al.," or "(Brants, 2000)."
   (c) Creation of text span: A text span is generated from the reference paper (RP) to include the most relevant sentences related to the citations from its citing papers (CP). However, for generic text summarization, such as news (DUC-2001 and DUC-2002) and scientific documents without citations, the entire document is considered as the text span.
   (d) Noise removal and Lower-casing: Noises such as HTML tags, punctuation marks, hyperlinks, and stop words are removed from the sentences. Additionally, the sentences are converted to lowercase to ensure consistency in further processing.

   These pre-processing steps help prepare the dataset for subsequent stages of the summarization process.

2. Population initialization: In the population initialization stage of our proposed system, each potential summary, referred to as an individual, is represented by a binary vector of length N, where N corresponds to the number of sentences in the text span. The binary

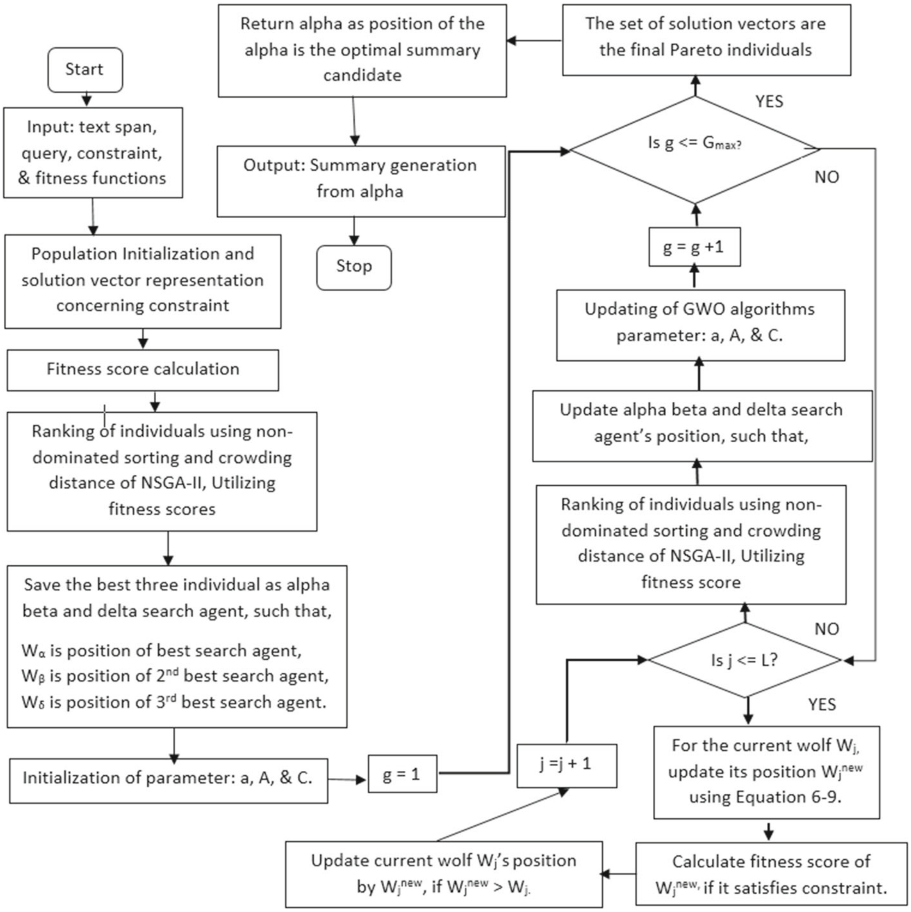

Start
Input: text span, query, constraint, & fitness functions
Population Initialization and solution vector representation concerning constraint
Fitness score calculation
Ranking of individuals using non-dominated sorting and crowding distance of NSGA-II, Utilizing fitness scores
Save the best three individual as alpha beta and delta search agent, such that,
Wα is position of best search agent,
Wβ is position of 2nd best search agent,
Wδ is position of 3rd best search agent.
Initialization of parameter: a, A, & C.
g = 1
Is g <= Gmax?
YES
NO
The set of solution vectors are the final Pareto individuals
Return alpha as position of the alpha is the optimal summary candidate
Output: Summary generation from alpha
Stop
g = g +1
Updating of GWO algorithms parameter: a, A, & C.
Update alpha beta and delta search agent's position, such that,
Ranking of individuals using non-dominated sorting and crowding distance of NSGA-II, Utilizing fitness score
Is j <= L?
NO
YES
For the current wolf Wj, update its position Wjnew using Equation 6-9.
Calculate fitness score of Wjnew, if it satisfies constraint.
Update current wolf Wj's position by Wjnew, if Wjnew > Wj.
j =j + 1

Fig. 1 Flowchart of the proposed approach: Binary Grey Wolf Optimization based Scientific Document Summarization

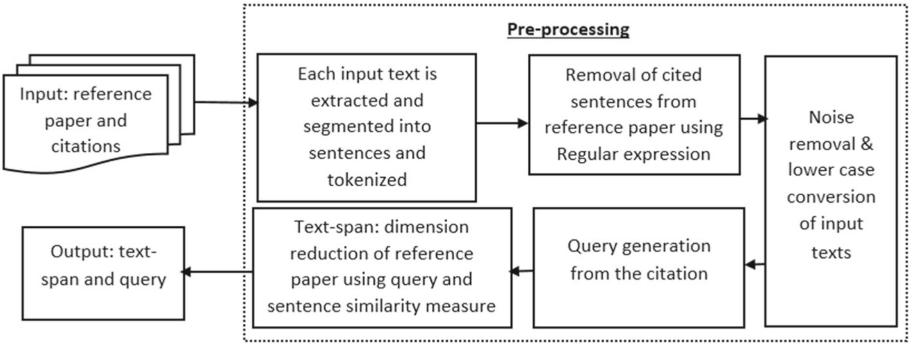

Pre-processing
Input: reference paper and citations
Each input text is extracted and segmented into sentences and tokenized
Removal of cited sentences from reference paper using Regular expression
Noise removal & lower case conversion of input texts
Query generation from the citation
Text-span: dimension reduction of reference paper using query and sentence similarity measure
Output: text-span and query

Fig. 2 Pre-processing of scientific documents

vector consists of sequences of zeros and ones, with a value of 1 indicating the presence of a sentence and 0 indicating its absence. These sentences are encoded sequentially based on their order in the text span. For example, a vector [1, 0, 0, 0, 1] represents the presence of the first and fifth sentences in the encoded individual, while the document has a total of five sentences.

To ensure feasibility, a constraint is applied based on the desired summary length. In this paper, we have used a constraint of 250 words, indicating that the summary length limit is set to 250 words. The summation of the words in a summary is evaluated against this constraint, and if it does not exceed the specified limit, the individual is considered feasible. From the pool of randomly generated feasible individuals, a subset is then selected to form the initial population for subsequent processing and optimization.

3. Fitness function formulation: Each individual's fitness is measured using the objective function score. In this paper, we have used three systems (with varying objective function formulations) for identifying the best among them, concerning news documents. In the first system, **TS-BGWO1objective**, the objective function score for an individual is calculated as the sum of all the feature scores. For example, an individual has five sentences, and six features represent each sentence. Then, the six-feature score for each sentence is calculated and then summed up to get the objective/fitness score for the individual. In the second system, **TS-BGWO6objective**, the objective function score is calculated considering each feature as an objective. In the third system, **TS-BGWO4objective**, four objective functions are considered. The first objective function score is calculated considering the sum of the normalized (scaled) score of (F_1, F_2, and F_3) features that ensure informativeness. F_4, F_5, and F_6 are considered the second, third, and fourth objective functions, respectively ensuring content coverage, anti-redundancy, and cohesion. Note that the objective function scores are calculated after checking the feasibility. If an individual is not feasible, its objective function score is zero. All these features are described in detail in Section 2.3.
   Similarly, for the citation-based SDS system, three systems are designed; (i). In the first system, **Sci-TS-BGWO1objective**, one fitness function is designed for which eight features are used. The sum of the eight feature scores is considered one fitness function score. (ii). In the second system, **Sci-TS-BGWO8objective**, each feature is considered an objective, and hence, eight fitness functions are designed. (iii). In the third system, **Sci-TS-BGWO5objective** , aspect-based objective functions are designed. For this purpose, each aspect is considered an objective, and in total five functions are designed. These objectives are: the normalized (scaled) informative features score's summation (F_1, F_2, and F_3); content coverage (F_4), anti-redundancy (F_5), cohesion (F_6), and the normalized (scaled) citation similarity score's summation (F_7, and F_8). Hence, in the third system, five fitness functions are used. Three more systems are designed with similar fitness functions but without pre-processing of reference paper to show the importance of pre-processing. Also, by taking the whole document as text span (without citations) some systems can be designed.
4. Wolf population distribution and ranking: Based on the objective function score, the wolf population is sorted using non-dominated sorting and crowding distance. The best wolf, P[0], is then classed as a "alpha wolf," followed by the second-best wolf, P[1], the third-best wolf, P[2], and the remaining wolf is considered as "omega wolf."
5. Position updation: All the wolf's position is updated in each iteration. Mathematically following equations are used for each wolf to update their position. As a binary

vector describes each wolf's position, each bit is updated during position updating using the (9).

$$(W_j^{new}) = crossover(Y_1^d, Y_2^d, Y_3^d) = \begin{cases} Y_1^d, & \text{if } R_6 < \frac{1}{3} \\ Y_2^d, & \text{if} \frac{1}{3} \leq r_6 < \frac{2}{3} \\ Y_3^d, & \text{otherwise} \end{cases} \tag{9}$$

where, Variable d is the dimension of search space, and R_6 is a random number uniformly distributed between [0,1] [35]. Y_1^d, Y_2^d, and Y_3^d are the binary vectors created by moving the alpha, beta, and delta wolves, respectively. These vectors are formed using the following (10), (11) and (12).

$$Y_1^d = \begin{cases} 1, & \text{if} (W_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Where W_α^d is the current position of the α wolf and $bstep_\alpha^d$ is a random move of the present wolf towards α.

$$Y_2^d = \begin{cases} 1, & \text{if } (W_\beta^d + bstep_\beta^d) \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Similarly, W_β^d is the current position of the β wolf and $bstep_\beta^d$ is a random move of the current wolf towards β wolf.

$$Y_3^d = \begin{cases} 1, & \text{if } (W_\delta^d + bstep_\delta^d) \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

And, W_δ^d is the current position of the δ wolf and $bstep_\delta^d$ is a random move of the current wolf towards δ wolf.
Suppose the new position of the wolf is feasible. In that case, the fitness of both the current and new positions is calculated, and the best position is considered as the wolf's updated position.

6. Iteration: In the proposed approach, the procedure of individuals position updation and Ranking is repeated until a maximum number of generations or iteration (G_{max}), where G_{max} is a user given input as shown in the Algorithm 1. However, in this system, if the α, β, and δ do not update their position in three consecutive iterations, the iteration stops without executing the further iterations, which further helps to minimize the computational cost.
7. Summary generation: The α wolf is considered the optimal wolf after the last iteration. Its corresponding summary is generated from the clean actual sentences in chronological order.

Algorithm 1: The pseudo-code of binary grey wolf optimization.

```
Input: Single Document
Output: The optimal summary
Initialize initial grey-wolf population $P[L][N]$, such that, number of individual
= $|L|$, length of individual as = $|N|$, maximum number of iteration = $G_{max}$.
Calculate fitness functions score of the individuals.
Sort the population P using non-dominated sorting and crowding distance of
NSGA-II utilizing their fitness scores.
Set:
  – $W_\alpha$= the position of best wolf.
  – $W_\beta$=The position of second best wolf.
  – $W_\delta$=the position of third best wolf.
for $i$ in range (1 to $G_{max}$) do
    for $j$ in range( 1 to $L$) do
        Compute $Y_1^d$, $Y_2^d$, and $Y_3^d$ for $W_j$ using (10), (11), and (12).
        Generate ($W_j^{new}$) by applying crossover between ($Y_1^d$, $Y_2^d$, $Y_3^d$) using (9).
        Calculate objective function score of ($W_j^{new}$), if ($W_j^{new}$) satisfies the
        constraints.
        Update $W_j$ by ($W_j^{new}$) if objective function score of ($W_j^{new}$) is $\geq$ objective
        function score of $W_j$.
    end for
    Sort P using non-dominated sorting and crowding distance and update the
    position of $\alpha$, $\beta$, $\delta$, and $\omega$.
end for
Return $\alpha$ as the optimal wolf and its corresponding summary as a system-generated
summary.
```

4 Experimental setup and result discussion

In this section, we provide a brief description of the experimental setup and discuss the results obtained from our proposed evolutionary approach for text summarization.

4.1 Data-set and evolutionary measures used

We have used open-source DUC[1] (Document Understanding Conferences) Corpus, namely, DUC 2001, DUC 2002. These data sets are clustered into 30 & 59 topics, containing 309 and 567 document-summary pairs, respectively. Details of this corpus is described in [10, 32]. We have also used the CL-SciSumm 2018 test data-sets[18][2] containing twenty reference papers, their citation and gold summaries. Scisummnet data-set[3] to show that the proposed system is also able to produce summaries on the large, structured dataset. The ScisummNet

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

[2] https://github.com/WING-NUS/scisumm-corpus

[3] https://cs.stanford.edu/~myasu/projects/scisumm_net

dataset contains the 1000 most cited papers in the ACL Anthology Network (AAN), with their annotated citation information and gold summaries.

To evaluate the quality of the generated summaries, we employed the ROUGE-1 and ROUGE-2 measures. These measures assess the similarity between the generated summaries and the human-created reference summaries by counting the number of overlapping units, such as n-grams, word sequences, and word pairs. Specifically, we focused on the recall scores of ROUGE-1 and ROUGE-2. Recall scores indicate the proportion of correct words identified by the generated summaries compared to the words in the reference summaries.

4.2 Comparing methods

For comparison, existing approaches concerning DUC-2001 and DUC-2002 dataset, namely, ESDS-AMGA-2 [10], ESDocSum [31], ESDS-SMODE [32], COSUM [2], and ETS-GA [36] are used. On ScisummNet data-set only few researchers have reported their work. So, for comparison prpose we have used $CL - SMODE_{WMD}$ [9], Continues BERT [29], and BERTSUM [25]

4.3 Result obtained on DUC-2001 and DUC-2002 dataset

Table 1 presents the ROUGE-1 and ROUGE-2 scores achieved by our proposed system and other existing systems. The corresponding scores are visualized in Fig. 3. To determine the relative rankings of these methods, we applied the ROUGE score-based Unified Ranking method, which has been used in previous studies [1, 11, 31]. The ranking calculation is performed using (13).

In (13), M represents the total number of methods compared, while R_p indicates the number of times a method appears at the p^{th} position. Based on the calculated $Ranking_{score}$ using (13), the methods are ranked. In our case, we have a total of eight methods, each with four values (ROUGE-1 and ROUGE-2 scores on the DUC-2001 and DUC-2002 datasets). Using (13), we calculate the $ranking_{scores}$ for each method and present them in Table 2. These scores are then sorted in descending order to determine the final rankings of the methods.

$$Ranking_{score} = \sum_{i=1}^{M} \frac{(M - p + 1)R_p}{M} \tag{13}$$

Table 1 The ROUGE-1 and ROUGE-2 scores obtained by different systems

Method Name	DUC-2001		DUC-2002	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
TS-BGWO4objective	0.53333	0.32059	0.54178	0.34256
TS-BGWO6objective	0.50956	0.30235	0.52132	0.31856
TS-BGWO1objective	0.48715	0.28542	0.50039	0.28506
ESDS-AMGA2 [10]	0.507692	0.295065	0.525813	0.312867
ESDocSum [31]	0.50236	0.29238	0.51662	0.28846
ESDS-SMODE [32]	0.45214	0.2145	0.49117	0.34132
COSUM [2]	0.4727	0.2012	0.4908	0.2309
ETS-GA [36]	0.45058	0.19619	0.48423	0.22471

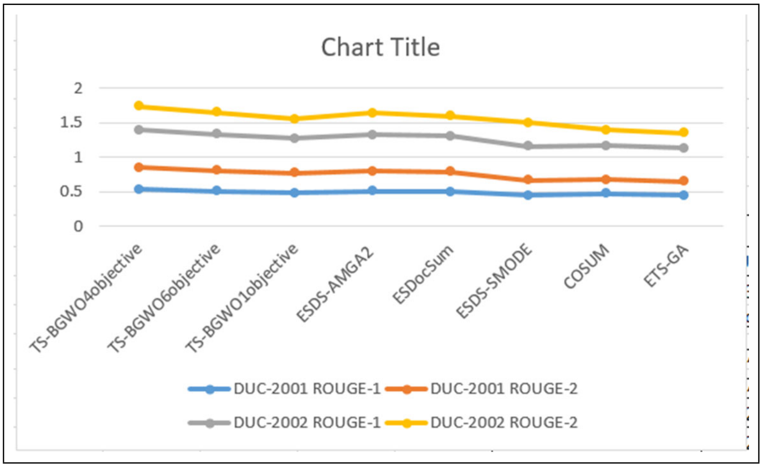


Fig. 3 Graph showing the ROUGE scores obtained by different approaches on DUC-2001 and DUC-2002 data-sets

Based on the information presented in Table 1, Fig. 3, and Table 2, it can be observed that the system TS-BGWO4objective outperforms all the existing systems mentioned in the paper. ESDS-SMODE [32] achieves a comparable ROUGE-2 score on the DUC-2002 dataset. The second best performing system is TS-BGWO6objective. However, the system utilizing a single objective function (TS-BGWO1objective) performs relatively poorly in comparison.

4.4 Study on effectiveness of objective function formulation on DUC-2001 and DUC-2002 datasets

In the TS-BGWO4objective system, four objective functions are formulated to address the key aspects of summarization: informativeness, content coverage, anti-redundancy, and cohesion. These aspects ensure that the generated summaries are relevant and readable. On the other hand, the TS-BGWO6objective system assigns equal importance to each feature, resulting in

Table 2 Ranking of different methods based on their ROUGE-1 and ROUGE-2 scores

Method Name	Rp								Ranking score	Rank
	1	2	3	4	5	6	7	8		
TS-BGWO4objective	4								4	1
TS-BGWO6objective		2	2						3.25	2
TS-BGWO1objective					3	1			1.875	5
ESDS-AMGA2 [10]		1	2	1					3	3
ESDocSum [31]				3	1				2.375	4
ESDS-SMODE [32]		1				2	1		1.875	6
COSUM [2]						1	3		1.125	7
ETS-GA [36]								4	0.5	8

R_p denotes how many times a method appeared at the p^{th} position, where p = 1,2,...8

six objective functions that are further optimized. In contrast, the TS-BGWO1objective system combines all these features into a single objective function, where the score is calculated as the sum of the scores for each feature. These objective functions are evaluated using the DUC-2001 and DUC-2002 datasets. Based on the ROUGE scores, the TS-BGWO4objective system demonstrates the best performance among the proposed systems, followed by the TS-BGWO6objective system in the second position, and the TS-BGWO1objective system in the third position.

Among the proposed and existing systems discussed in this paper, the ESDS-AMGA2 system ranks third. This system is based on a multi-objective framework that utilizes an archive-based micro genetic-2 algorithm, resulting in effective performance. However, our proposed system differs in terms of the feature set and the grey wolf optimization algorithm used. The fourth-ranked system is ESDocSum, which is similar to our second-best system (TS-BGWO6objective). ESDocSum employs six objective functions, including sentence position, title similarity, sentence length, cohesion, coverage, and readability factor. These features differ from our own feature set in ESDocSum. Additionally, the underlying strategy in ESDocSum is distinct. ESDS-SMODE and COSUM adopt an optimization strategy to determine the number of clusters and select sentences from these clusters using specific sentence selection techniques. On the other hand, ETS-GA employs a genetic algorithm-based single objective optimization approach. After careful analysis, we conclude that our proposed system offers a minimal computational cost while producing readable and relevant summaries.

4.5 Statistical significance

We conducted a One-way ANOVA test to determine the significance of the proposed system. Seven runs were performed for each of the three proposed systems, and their ROUGE scores were obtained by slightly varying them for the existing systems. Table 3 presents the parameters used, and Table 4 displays the obtained p-values. Two hypotheses were considered: the null hypothesis stating that there are no significant differences in the scores between the proposed and existing methods, and the alternative hypothesis suggesting that significant changes exist. The average ROUGE-1 recall scores on the DUC-2001 dataset were taken first, with some variation based on the scores reported in the respective papers for the existing systems. Similarly, the ROUGE-2 recall score for DUC-2001, as well as the ROUGE-1 and ROUGE-2 scores for DUC-2002, were included. The p-value obtained from Table 4, corresponding to the F-statistic of the one-way ANOVA test, was found to be lower than 0.05. This indicates that one or more systems show significant differences, leading to the rejection of the null hypothesis.

However, to assess the significant differences between the systems, we conducted the Tukey HSD Test instead of the One-way ANOVA test. The critical values used were as follows: the number of methods (m) was 8, the degrees of freedom for the error term (ν) were 46, and the significance levels (α) were set to 0.01 and 0.05. Using these critical values, we obtained Q values: for $\alpha = 0.01$, $Q_{criticle}^{\alpha=0.01,m=8,\nu=46} = 5.3369$ and $Q_{criticle}^{\alpha=0.05,m=8,\nu=46} = 4.4893$, respectively. Tukey-Kramer HSD -statistic is calculated using the (14).

$$Q_{i,j} = \frac{|\bar{x_i} - \bar{x_j}|}{S_{i,j}} \tag{14}$$

Table 3 One-way ANOVA test parameters, tested using ROUGE-1 scores of DUC-2001 data-set

Systems	A	B	C	D	E	F	G	H	Total
observations N	7	7	7	7	7	7	7	7	56
sum	2.6798	3.5869	3.4201	3.5566	3.5304	3.1669	3.3089	3.1541	26.404
mean	0.536	0.5124	0.4886	0.5081	0.5043	0.4524	0.4727	0.4506	0.489
sum of squares	1.4365	1.8381	1.671	1.8071	1.7806	1.4327	1.5641	1.4212	12.951
sample variance	0.0001	0	0	0	0	0	0	0	0.0008
sample std. dev.	0.0079	0.0049	0.0038	0.0009	0.003	0.0004	0	0	0.0279
std. dev. of mean	0.0035	0.0018	0.0014	0.0004	0.0011	0.0001	0	0	0.0038

Eight systems are A(TS-BGWO4objective), B (TS-BGWO6objective), C(TS-BGWO1objective), D(ESDS-AMGA2), E(ESDocSum), F(ESDS-SMODE), G(COSUM), and H(ETS-GA)

Table 4 One-way ANOVA test results

source	sum of squares	degrees of freedom	mean square MS	F statistic	p-value
methods	0.0406	7	0.0058	492.92	1.1102e-16
error	0.0005	46	0		
total	0.0412	53			

The denominator in the above expression is calculated using (15). In this equation, $H_{i,j}$ represents the harmonic mean of the number of observations in the columns labeled as i and j, and $\hat{\sigma}_{\epsilon}$ represents the square root of the Mean Square Error.

$$S_{i,j} = \frac{\hat{\sigma}_{\epsilon}}{\sqrt{H_{i,j}}}; i, j = 1, 2, ...k; i \neq j \tag{15}$$

The Table presents the Tukey HSD Test results of the evaluation, whether $Q_{i,j} > Q_{criticle}$ for all relevant pairs of treatments. In addition, we also present the significance (p-value) of the observed. Q-statistic. All the obtained values respective to the system TS-BGWO4objective are 0.05% significant, more specifically, 0.01 per cent effective. A similar approach has been used for testing the other obtained ROUGE values (ROUGE-2 score for DUC-2001 and ROUGE-1 and ROUGE-2 score for DUC-2002 dataset). All these values are also statistically significant.

4.6 Study on effectiveness of objective function formulation on scientific document summarization

After analyzing the results of the proposed system on DUC-2001 and DUC-2002, similar systems were built to evaluate the system's performance on the ScisummNet dataset using the same approach. The objective of these experiments was to assess the suitability of the approach for large single scientific documents, particularly in the context of citation-based scientific document summarization. Scientific documents differ from news documents in several aspects, such as length and structure. Therefore, this paper primarily focused on reducing the length of the documents, specifically by formulating summaries based on the text spans of referenced papers. The following experiments were conducted and included in the paper:

1. Importance of pre-processing scientific documents: Systems were designed with and without pre-processing to demonstrate the impact of pre-processing on the summarization results.
2. Fitness function formulation: Three systems were designed to showcase different formulations of the fitness function.
3. Summary generation without citation utilization: A summary was created using the entire document without considering citations, highlighting the advantages of citation-based generation.
4. Comparative analysis of ROUGE scores: A comparative analysis of the ROUGE scores was conducted to evaluate the performance of the proposed system in comparison to existing scientific document systems.

These experiments aimed to provide insights into the effectiveness of the proposed approach for scientific document summarization, considering the unique characteristics of scientific documents and the utilization of citations. These systems are briefly presented in Table 5. In $Sci - TS - BGWO1objective$, the objective function is formulated by taking the sum of all eight features; hence, it is a single objective system. The pre-processing and creation of text-span concerning citation contextualization is performed in this system. In $Sci - TS - BGWO5objective$, five objectives are taken. Where four objectives are similar to the system ($TS - BGWO4objective$), and one more objective (citation contextualization) is formulated considering F_7, and F_8 feature. Each of the eight features is considered as one fitness function or objective in $Sci - TS - BGWO8objective$, hence, eight objective functions. The systems $Sci - TS - BGWO1objective - NTR$, $Sci - TS - BGWO4objective - NTR$, and $Sci - TS - BGWO6objective - NTR$ are generic systems where the whole scientific document is considered for summarization. In $Sci - TS - BGWO1objective - NPP$, the objective function is formulated by taking the sum of all the eight features; hence, it is a single objective system, but pre-processing concerning citation contextualization is not performed. Similarly, pre-processing concerning citation contextualization is not performed in $Sci - TS - BGWO5objective - NPP$, and $Sci - TS - BGWO8objective - NPP$. In all the multi-objective systems, we have used the non-dominated sorting and crowding distance of NSGA-II for sorting the individuals. Table 6, and Fig. 4 show the results that were achieved. From the Table 6, and Fig. 4, it is visible that $Sci - TS - BGWO5objective$ outperformed these systems. Here we have formulated aspect-based objectives, i.e., five objectives: content coverage, informativeness, anti-redundancy, cohesion, and citation contextualization. The second-best system is $Sci - TS - BGWO4objective - NTR$, where the whole document is considered for summary generation, and the third-best system is $Sci - TS - BGWO8objective$, where all the objectives are given equal weight and the

Table 5 System' Description concerning ScisummNet data-set

System Name	Fitness Function(FF) Used	Pre-Processing(PP)/ text reduncion (TR) concering citations citextualization
$Sci - TS - BGWO1objective - NPP$	1 FF : sum of 8 features	TR, no PP
$Sci - TS - BGWO5objective - NPP$	5 FF: five aspects.	TR, no PP
$Sci - TS - BGWO8objective - NPP$	8 FF: 8 feature as 8 objective	TR, no PP
$Sci - TS - BGWO1objective - NTR$	1 FF : sum of 8 features	PP, no TR
$Sci - TS - BGWO4objective - NTR$	4 FF: four aspects.	PP, no TR
$Sci - TS - BGWO6objective - NTR$	6 FF: 6 feature as 6 objective	PP, no TR
$Sci - TS - BGWO1objective$	1 FF : sum of 8 features	PP and TR
$Sci - TS - BGWO5objective$	5 FF: five aspects.	PP and TR
$Sci - TS - BGWO8objective$	8 FF: 8 feature as 8 objective	PP and TR

Table 6 System's performance in summary generation, on ScisummNet data-set

System Name	ROUGE-1 Recall	ROUGE-1 F1-measure	ROUGE-2 Recall	ROUGE-2 F1-measure
$Sci - TS - BGWO1objective-NPP$	35.3	34.2	13.8	12.4
$Sci - TS - BGWO5objective-NPP$	39.8	38.2	16.5	14.3
$Sci - TS - BGWO8objective-NPP$	38.8	37.9	16.1	13.9
$Sci - TS - BGWO1objective-NTR$	37.5	35.4	14.9	13.2
$Sci - TS - BGWO4objective-NTR$	46.2	43.6	25.9	27.5
$Sci - TS - BGWO6objective-NTR$	42.8	40.5	22.8	19.5
$Sci - TS - BGWO1objective$	40.3	38.9	18.3	17.9
$Sci - TS - BGWO5objective$	48.4	44.5	27.2	28.9
$Sci - TS - BGWO8objective$	45.9	42.3	23.6	21.3

system is citation contextualization-based. The less-performing systems are single-objective function-based systems. From the Table 6, it is visible that using proper pre-processing improved the result. However, citation contextualized systems, including pre-processing, are the best-performing system.

Proposed systems F1-scores are compared with existing systems and presented in the Table 7. We have only used F1-scores as these system has provided only the F1-scores. From

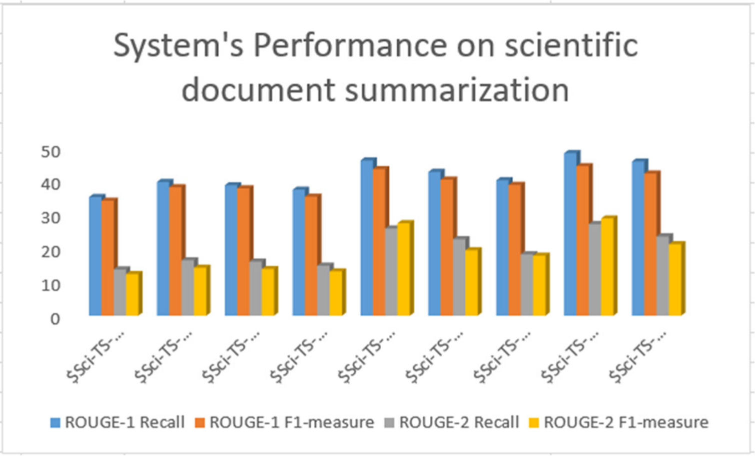


Fig. 4 System's performance on scientific document summarization

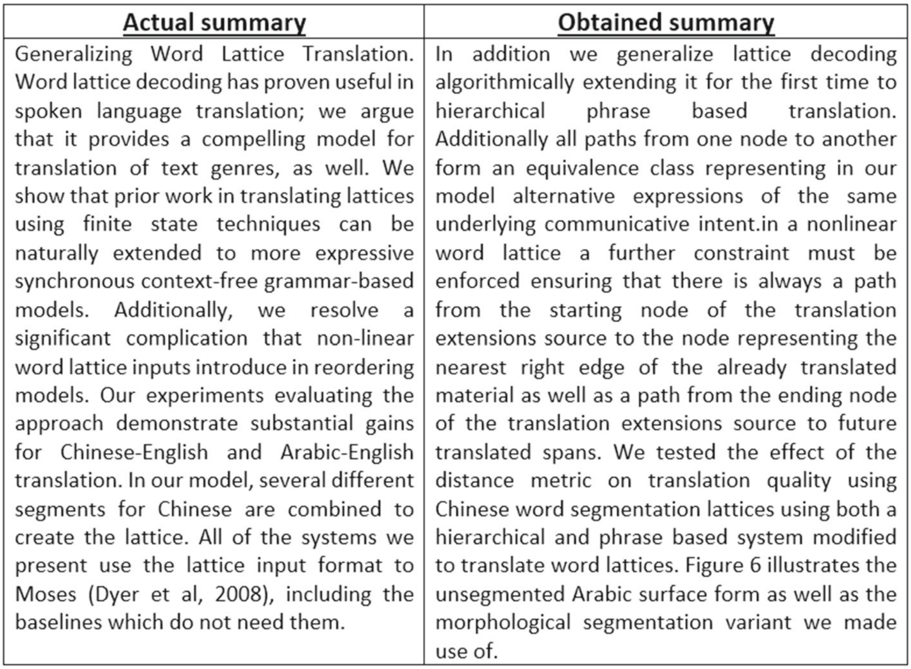
Table 7 Proposed and existing systems performance on ScisummNet data-set

System Name	ROUGE-1 F	ROUGE-2 F
Continual BERT	31.6	12.7
BERTSUM	33	13.4
$CL - SMODE_{WMD}$	43.5	23.9
Sci-TS-BGWO5objective	44.5	28.9

the Table 7, it is visible that the proposed system have outperformed all the existing systems, mentioned in this paper.

4.7 Obtained summary analysis

After analyzing selected summaries from each of the datasets, including the ScisummNet dataset, it was observed that the obtained summaries were semantically similar to the actual summaries, despite having low ROUGE scores. Additionally, all the obtained summaries were found to be readable and relevant to their respective documents. One example of such a summary from the ScisummNet dataset is depicted in Fig. 5. These findings suggest that the proposed system was able to generate summaries that captured the essence of the original documents, even if the ROUGE scores alone might not fully capture their quality.



Actual summary	Obtained summary
Generalizing Word Lattice Translation. Word lattice decoding has proven useful in spoken language translation; we argue that it provides a compelling model for translation of text genres, as well. We show that prior work in translating lattices using finite state techniques can be naturally extended to more expressive synchronous context-free grammar-based models. Additionally, we resolve a significant complication that non-linear word lattice inputs introduce in reordering models. Our experiments evaluating the approach demonstrate substantial gains for Chinese-English and Arabic-English translation. In our model, several different segments for Chinese are combined to create the lattice. All of the systems we present use the lattice input format to Moses (Dyer et al, 2008), including the baselines which do not need them.	In addition we generalize lattice decoding algorithmically extending it for the first time to hierarchical phrase based translation. Additionally all paths from one node to another form an equivalence class representing in our model alternative expressions of the same underlying communicative intent.in a nonlinear word lattice a further constraint must be enforced ensuring that there is always a path from the starting node of the translation extensions source to the node representing the nearest right edge of the already translated material as well as a path from the ending node of the translation extensions source to future translated spans. We tested the effect of the distance metric on translation quality using Chinese word segmentation lattices using both a hierarchical and phrase based system modified to translate word lattices. Figure 6 illustrates the unsegmented Arabic surface form as well as the morphological segmentation variant we made use of.

Fig. 5 System generated and actual summary of the document P08-1115.xml from ScisumNet data-set

4.8 Complexity analysis

System Specification: The proposed system is designed, and evaluated using Lenovo legion laptop, having Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz 2.50 GHz processor, 8.00 GB Ram, 64-bit operating system, and x64-based processor.

Worst case complexity of TS-BGWO4objective: Let the size of the initial population, parent, and offspring in each iteration be N. Archive size (used for child generation to store α, β, and δ wolf) is 3, and hence, mating pool size for each individual is $3 + 1$. Let the total number of function evaluations be T, and the number of objectives is (M = 4).

Now, generation, and formation & updation of archive has complexity of $\theta(N)$ in worst case. Evaluation of population, takes $\theta(M * N^2)$ time. The sorting time taken is $\theta(MNlogN)$. Hence, per iteration, complexity of TS-BGWO4objective is $\theta(N^2logN)$ assuming $logN > M$ and the overall complexity is $\theta(TN^2logN)$ [10].

5 Conclusion

In this study, we demonstrate the effectiveness of a multi-objective extractive text summarization system suitable for single-text documents, including news, scientific, single-lingual, low- and high-resource documents. The system, based on binary Grey Wolf Optimization, encompasses various functionalities such as data pre-processing, feature scaling, similarity measurement, feature selection, fitness function formulation, position updating strategy, and automatic summary generation. We evaluate the system on the DUC-2001, DUC-2002, and ScisummNet datasets, achieving promising results.

The highest-performing system for the DUC-2001 and DUC-2002 datasets is $TS-BGWO4objective$, with ROUGE-1 scores of 0.53333 and 0.32059 on DUC-2001, and ROUGE-1 scores of 0.5417 and 0.34256 on DUC-2002. For scientific documents (ScisummNet datasets), the top-performing system is $Sci-TS-BGWO5objective$, achieving ROUGE-1 scores of 48.4 and ROUGE-2 scores of 27.2. These results indicate that our proposed system outperforms several state-of-the-art methodologies in terms of ROUGE scores. While the current study focuses on single-document summarization, our future work will explore modifications to enable multi-document summarization.

Funding no funds, grants, or other support was received.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest

References

1. Aliguliyev MR (2009) A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Exp Syst Appl 36(4):7764–7772
2. Aliguliyev RM, Aliguliyev RM, Isazade NR, Abdi A, Idris N (2019) COSUM: Text summarization based on clustering and optimization. Exp Syst 36(1):e12340
3. Assent I, Wichterich M, Meisen T, Seidl T (2008) Efficient similarity search using the earth mover's distance for large multimedia databases. In: 2008 IEEE 24th International conference on data engineering. pp 307–316. IEEE
4. Beltagy I, Peters ME, Cohan A (2020) Longformer: The long-document transformer. arXiv:2004.05150

5. Cao Z, Li W, Wu D (2016) Polyu at cl-scisumm 2016. In: Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL). pp 132–138
6. Chakraborty T, Narayanam R (2016) All fingers are not equal: Intensity of references in scientific articles. arXiv:1609.00081
7. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
8. Debnath D, Achom A, Pakray P (2018) NLP-NITMZ@ CLScisumm-18. In: BIRNDL@ SIGIR. pp 164–171
9. Debnath D, Das R (2022) Automatic citation contextualization based scientific document summarization using multi-objective differential evolution. In: Advanced techniques for IoT Applications: Proceedings of EAIT 2020. pp 289–301. Springer
10. Debnath D, Das R, Pakray P (2020) extractive single document summarization using an archive-Based Micro Genetic-2. In: 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI). pp 244–248. IEEE
11. Debnath D, Das R, Pakray P (2021) Extractive single document summarization using multi-objective modified cat swarm optimization approach: ESDS-MCSO. Neural Computing and Applications pp 1–16
12. Deng Z, Zeng Z, Gu W, Ji J, Hua B (2021) Automatic related work section generation by sentence extraction and reordering
13. Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381
14. Garzone M, Mercer RE (2000) Towards an automated citation classifier. In: Conference of the canadian society for computational studies of intelligence. pp 337–346. Springer
15. Ghodratnama S, Beheshti A, Zakershahrak M, Sobhanmanesh F (2020) Extractive document summarization based on dynamic feature space mapping. IEEE Access 8:139084–139095
16. Hernández-Alvarez M, Gomez JM (2016) Survey about citation context analysis: Tasks, techniques, and resources. Natural Lang Eng 22(3):327–349
17. Hong M, Wang H (2021) Research on customer opinion summarization using topic mining and deep neural network. Math Comput Simul 185:88–114
18. Jaidka K, Chandrasekaran MK, Rustagi S, Kan, MY (2016) Overview of the cl-scisumm 2016 shared task. In: Proceedings of the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL). pp 93–102
19. Jaidka K, Yasunaga M, Chandrasekaran MK, Radev D, Kan MY (2019) The cl-scisumm shared task 2018: Results and key insights. arXiv:1909.00764
20. Jana E, Uma V (2020) Opinion mining and product review summarization in E-Commerce. In: Trends and applications of text summarization techniques, pp 216–243. IGI Global
21. Khan A, Gul MA, Zareei M, Biswal R, Zeb A, Naeem M, Saeed Y, Salim N (2020) Movie review summarization using supervised learning and graph-based ranking algorithm. Computational intelligence and neuroscience 2020
22. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2019) Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461
23. Li L, Mao L, Zhang Y, Chi J, Huang T, Cong X, Peng H (2016) Cist system for cl-scisumm 2016 shared task. In: Proceedings of the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL). pp 156–167
24. Lin CY (2004) Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out. pp 74–81
25. Liu Y (2019) Fine-tune BERT for extractive summarization. arXiv:1903.10318
26. Ma S, Zhang H, Xu J, Zhang C (2018) Njust@ clscisumm-18. In: BIRNDL@ SIGIR
27. Miller GA (1995) WordNet: a lexical database for English. Commun ACM 38(11):39–41
28. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
29. Park JW (2020) Continual bert: Continual learning for adaptive extractive summarization of covid-19 literature. arXiv:2007.03405
30. Ramadhan MR, Endah SN, Mantau ABJ (2020) Implementation of Textrank Algorithm in Product Review Summarization. In: 2020 4th International Conference on Informatics and Computational Sciences (ICICoS). pp 1–5. IEEE
31. Saini N, Saha S, Chakraborty D, Bhattacharyya P (2019) Extractive single document summarization using binary differential evolution: Optimization of different sentence quality measures. PloS one 14(11):e0223477

32. Saini N, Saha S, Jangra A, Bhattacharyya P (2019) Extractive single document summarization using multi-objective optimization: Exploring self-organized differential evolution, grey wolf optimizer and water cycle algorithm. Knowl Based Syst 164:45–67
33. Sanchez-Gomez JM, Vega-Rodríguez MA, Pérez CJ (2020) A decomposition-based multi-objective optimization approach for extractive multi-document text summarization. Appl Soft Comput 91:106231
34. Tadashi N (2016) NEAL: A neurally enhanced approach to linking citation and reference. In: Proceedings of the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL). pp 168–174
35. Too J, Abdullah AR, Mohd Saad N, Mohd Ali N, Tee W (2018) A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification. Comput 7(4):58
36. Vázquez E, Arnulfo Garcia-Hernandez R, Ledeneva Y (2018) Sentence features relevance for extractive text summarization using genetic algorithms. J Intell Fuzzy Syst 35(1):353–365
37. Vladimir B (2003) Efficient algorithms for citation network analysis. arXiv:cs/0309023
38. Wang P, Li S, Wang T, Zhou H, Tang J (2018) Nudt@ clscisumm-18. In: BIRNDL@ SIGIR
39. Wan X, Peng Y (2005) The earth mover's distance as a semantic measure for document similarity. In: Proceedings of the 14th ACM international conference on Information and knowledge management. pp 301–302
40. Zerva C, Nghiem MQ, Nguyen NT, Ananiadou S (2020) Cited text span identification for scientific summarisation using pre-trained encoders. Scientometrics 125(3):3109–3137

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ranjita Das[1,2] · Dipanwita Debnath[2,3] · Partha Pakray[4] · Naga Chaitanya Kumar[2]

Dipanwita Debnath
ddebnath.nita@gmail.com

Partha Pakray
partha@cse.nits.ac.in

Naga Chaitanya Kumar
cnck1999@gmail.com

[1] Department of Computer Science & Engineering, National Institute of Technology Agartala, 799046 Tripura, India

[2] Department of Computer Science & Engineering, National Institute of Technology Mizoram, 796012 , Mizoram, India

[3] Department of CSE, Koneru Lakshmaiah Education Foundation, 522302 Vaddeswaram, India

[4] Department of Computer Science & Engineering, National Institute of Technology Silchar, 788010 Assam, India