e-palli

Journal of Natural Language and Linguistics (JNLL)

ISSN: 2995-9837 (ONLINE)

VOLUME 3 ISSUE 1 (2025)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Journal of Natural Language and Linguistics (JNLL)
Volume 3 Issue 1, Year 2025
ISSN: 2995-9837 (Online)
DOI: https://doi.org/10.54536/jnll.v3i1.3079
https://journals.e-palli.com/home/index.php/jnll

Text Summarisation Using Swarm-Based and Genetic-Based Methods in Natural Language Processing: A Comparative Review

Albaraa Abuobeida[1*]

Article Information

Received: June 04, 2024

Accepted: July 09, 2024

Published: June 26, 2025

Keywords

Genetic-Based Methods, Natural Language Processing, Particle Swarm Optimization, Swarm-Based Methods, Text Summarization

ABSTRACT

The digital age has resulted in unparalleled data availability, posing a serious problem of information overload. As a result, automatic text summarizing has evolved as an important technique for compressing large volumes of data into concise, consumable summaries such as Swarm-Based and Genetic-Based approaches. This comparative review investigates the use of those two approaches in text summarization within Natural Language Processing (NLP). Using an analysis of twelve studies, this study examines various techniques' types, benefits, drawbacks, and performance. The review concludes that swarm-based approaches, such as particle swarm optimization and ant colony optimization, excel in efficient search space exploration. Genetic-based techniques, such as genetic algorithms, provide further advancements. The comparison study sheds light on each approach's unique benefits and drawbacks, offering researchers and NLP practitioners insightful information. The study concludes that swarm-based methods are efficient for rapid convergence and continuous optimization, while genetic-based methods offer flexibility but require more computational resources. The choice depends on the specific requirements of the text summarization task. Future investigations have to concentrate on tackling issues like scalability and dependability and investigating the possibilities of merging Swarm-Based and Genetic-Based techniques.

INTRODUCTION

With emerging technologies such as Artificial Intelligence and virtual spaces, many business sectors, educational sectors, and even healthcare sectors have revolutionized themselves towards this digital transformation. These changes have helped many sectors streamline their process, enhancing the business's overall productivity and increasing their customers' experience. AI-powered tools such as ChatGPT and DeepBrain AI are constantly used in business for data analysis, user services, and marketing, resulting in more organized and successful decision-making (Rane, 2023). Virtual spaces are used in educational sectors to provide different learning experiences, such as virtual classrooms and online platforms for resources (Manegre & Sabiri, 2022). Overall, the use of digital technology has increased in different sectors, enhancing the work process and paving the way for a more interconnected and digital future.

Natural Language Processing (NLP) is a branch of Artificial intelligence that explores computers' interaction with human language. NLP aims to help computers understand, interpret, and generate human language so that it is meaningful, understandable, and applicable (Khurana *et al.*, 2023). NLP consists of numerous operations, such as text understanding, generating languages, speech synthesis, and recognition. Text understanding tasks include semantic analysis, parsing, and entity recognition to assist computers in comprehending the significance and context of text (Basha *et al.*, 2023). Machine translation, text summarization, and conversation production are language generation tasks in which computers construct human-like text depending on user input. Speech recognition allows computers to turn spoken language into text, whereas speech synthesis enables them to synthesize spoken language from text, allowing human-computer communication via speech (Vashisht *et al.*, 2021).

Researchers require information and resources as a reference to conduct comparable research, increasing the applicability of the findings and contributing to the growth of knowledge in their sector. However, a huge amount of data is present on the internet, which underlines the need for a summarization method, hence condensing the information while protecting the main idea of the research (Awasthi *et al.*, 2021). The concept of summarization helps to reduce the time spent reading. It enables faster information retrieval, enabling users to access extensive material on a certain topic more effectively.

Researchers have been summarizing information manually for a long time because it is a very effective process. By summarizing manually, it preserves the text's original meaning and reduces grammar mistakes overall. However, manual summarization consumes time and might produce inconsistent results (El-Kassas *et al.*, 2021). Furthermore, the method is subjective and can vary substantially according to the individual summarizer, which may result in a biased summary. Automated text summarization has appeared as a feasible alternative to address the limitations of manual summarizing. Automated summarizing uses computer algorithms to produce material summaries, making it more scalable and

[1] University of Hafr Al Batin, College of Computer Science and Engineering, Al Jamiah, Hafar Al Batin 39524, Saudi Arabia
[*] Corresponding author's e-mail: albarraa@uhb.edu.sa

efficient (Nazari & Mahdavi, 2019). However, automatic text summarization confronts several obstacles, including consistency, reliability, scalability, and fluency (Widyassari *et al.*, 2022). Evaluation metrics are also important for ensuring the accuracy of the summaries produced.

For text summarization, swarm-based and genetic-based methods are disparate approaches inspired by different principles. Swarm-based approaches can improve summarization, especially regarding feature or word selection (Rostami, Berahmand, Nasiri, *et al.*, 2021). Swarm-based methods include Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO), which provide the best solutions with the help of the collective behavior of decentralized systems. At the same time, the Genetic-Based methods, including the Genetic Algorithm (GA), use the natural selection process to generate solutions (Sohail, 2023).

This review examines the disparities between Swarm-Based and Genetic-Based methods for text summarization in Natural Language Processing (NLP). Although these methods have been used for different NLP tasks, much research hasn't been done on how well they compare when it comes to text summarization. This study clarifies the distinct advantages and disadvantages of Swarm-Based and Genetic-Based approaches. The swarm-based methods are preferable for efficiently searching the search space, even if the genetic-based techniques could provide a few improvements and additions. By offering insightful information about their effectiveness, this comparative analysis aims to progress the area of NLP and improve already existing summarizing methods.

LITERATURE REVIEW

Text Summarization in Natural Language Processing (NLP)

The growing numbers of internet users and the collection of enormous amounts of data online demand the usage of extractive summarization to condense this information while retaining its original meaning (El-Kassas *et al.*, 2021). Deep learning has transformed NLP by enabling access to large amounts of data and powerful computational power. Pre-trained transformer-based models, such as BERT and GPT, have dramatically increased effectiveness for text summarization tasks, yielding cutting-edge results (Yang *et al.*, 2023). Text summarization continues to be an issue in NLP and information retrieval, owing to the exponential expansion of digital content. Researchers have created artificial text summarizing algorithms that may be divided into two categories: extractive and abstractive summarization. Over the years, researchers refined and enhanced these strategies, utilizing developments in machine learning, deep learning, and natural language processing (Abualigah *et al.*, 2020). Extractive-based text summarization has numerous challenges, such as identifying the most relevant sentences from the source document (Abuobieda *et al.*, 2013).

The Gravitational Search Algorithm (GSA) is increasingly used in Natural Language Processing (NLP), notably for extractive summarization work (Mosa, 2020). GSA-based techniques concentrate on finding sentences with linguistic and statistical features to extract important information efficiently. The exponential rise of textual data, particularly in industries such as tourism, emphasizes the significance of NLP and analysis. NLP, which uses machine learning techniques, allows researchers to derive significant information from unstructured textual data, resulting in a better knowledge of social phenomena and more informed decision-making (Egger & Gokce, 2022).

Feature Selection in Text Summarization

A range of techniques, such as feature-based, graph-based approaches, and cluster-based, have been suggested to improve this feature selection procedure. A hybrid approach combining feature-based and cluster-based methods has generated high-quality single-document summaries. Rather than using the genetic mutation operator for real-to-integer values, it suggests a novel modulator based on the Jaccard correlation measure to modify sentence clustering (Abuobieda *et al.*, 2013).

In 2012, an article emphasized the relevance of feature selection in text summarization. It underlined that not all traits are equally significant, and it is recommended that we adopt a (pseudo) genetic idea to improve this process. The study sought to enhance the summary quality by randomly picking five features. To train the model and calculate feature weights, researchers used the Document Understanding Conference (DUC2002) dataset (Abuobieda *et al.*, 2012). The model showed potential in feature selection, resulting in higher summary quality. Text features are critical in recognizing significant concepts in a document for summary. Another study describes a feature selection strategy for extractive single-document summarization that uses a (pseudo) Genetic Probabilistic-based Summarization (PGPSum) model. The algorithm derives feature values from texts and optimizes feature scores to choose key lines for the summary. Experimental findings reveal that the PGPSum model surpasses benchmarks such as Ms-Word and Copernic summarizers, delivering summaries comparable to human-generated ones (Abuobieda *et al.*, 2011).

Feature selection is an important procedure in machine learning, especially given the exponential rise of high-dimensional datasets (Tsamardinos *et al.*, 2019). The problem of dimensionality, in which the quantity of characteristics greatly outnumbers the number of patterns, poses a substantial difficulty for applications that use data mining. Swarm Intelligence (SI) algorithms have become known as powerful tools for feature selection in big datasets. SI algorithms, influenced by collective actions in decentralized systems, can optimize feature selection by effectively exploring and utilizing the search space (Brezočnik *et al.*, 2018). This work presents a detailed evaluation of SI-based feature-selecting algorithms, grouping them according to their biological determination. The paper attempts to address an absence in understanding SI algorithms for feature selection by providing insights into cutting-edge methodologies and applications.

Cases in Summarisation Tasks Using NLP Methods

In a paper published in 2018, the authors suggested an automated text summary technique for court rulings. They underline the significance of conducting extensive research on legal issues and the difficulty of obtaining relevant information from long decisions. The system captures concepts inside single documents using a natural language processing approach known as Latent Semantic Analysis (LSA). Two techniques are utilized depending on the kind of input case (criminal or civil): single-document untrained and multi-document trained. The data for the system was gathered from official government websites, including Supreme Court, High Court, and District Court cases. The method received a median ROGUE-1 score of 0.58 and was authorized by competent attorneys (Merchant & Pande, 2018). Future developments will improve continuity within produced summaries and provide more reliable evaluations.

In 2020, the authors proposed a text summary system based on the BERT and GPT-2 models to solve the difficulty of summarizing the constantly developing COVID-19-related material. They use the COVID-19 Open Research Dataset to perform extractive summarization using BERT and abstract summarization with GPT-2 and evaluate the outcomes using ROUGE scores and visual examination. Their technique seeks to produce thorough summaries using keywords retrieved from original papers, allowing the medical community to obtain concise summaries of studies in which abstracts are not provided. Despite obstacles such as a lack of domain-specific corpora and obscure scientific language, the framework appears to be a promising option, enabling further growth and assisting in developing COVID-19 therapies and interventions (Kieuvongngam *et al.*, 2020).

Text summarization has grown in popularity as digital information increases, making it difficult for people to absorb vast quantities of text efficiently. Natural Language Processing (NLP) approaches have been essential in improving text summarizing methods, making it simpler to gather key information from text. Summarizing has become more accessible and successful because of techniques like extractive summarizing, which picks key sentences or phrases from the source material, and abstractive summarization, which provides summaries by paraphrasing the original text (Boorugu & Ramesh, 2020). These approaches employ NLP to break down text more intelligently, enhancing the quality and usability of summaries and allowing readers to rapidly understand the major ideas of a document without needing to read the full text.

MATERIALS AND METHODS

The methodology includes the search strategies below to help you better understand the articles. The use of a database collecting strategy has been very impactful in determining the overall use of the algorithms, as seen in Table 1:

To guarantee the accuracy of the review, certain inclusion and exclusion criteria that outline the selection procedure have been developed, as shown in Table 2:

Table 1: Search Strategies

S. no	Search Strategies
1.	("Text Summarization") AND ("Natural Language Processing") AND ("Swarm-Based Methods") AND (Genetic-Based Methods") AND ("Text Summarization using Algorithms") AND ("Summarizing Court Cases using NLP") AND ("Summarizing Medical Papers using NLP") AND ("Feature Selection in Text Summarization").
2.	("Text Summarization using Particle Swarm Optimization") AND ("Text Summarization using Ant Colony Optimization") AND ("Text Summarization using Genetic Algorithms") AND ("Text Summarization using Genetic Population") AND ("Text Summarization using Differential Evolution").
3.	("Comparison of Genetic-Based Methods with Swarm-Based Methods") AND ("Challenges of Genetic-Based Methods") and ("Challenges of Swarm-Based Methods") AND ("Performance of Swarm-Based Methods") AND ("Performance of Genetic-Based Methods").

Table 2: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies focusing on text summarization using swarm-based and genetic-based methods in Natural Language Processing (NLP).	Studies that do not focus on text summarization using swarm-based and genetic-based methods in Natural Language Processing (NLP).
Studies published within the last six years (2018-2023) have been included.	Studies older than six years have been excluded.
Studies that have used different types of swarm-based and genetic-based methods.	Studies that do not have used different types of swarm-based and genetic-based methods.
Studies related to the keywords in the title were included.	Studies not related to the keywords from the title were excluded.

Table 3: Article Selection

S. no	Author	Journal/Dissertation	Year	Findings
1.	Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, Hoda K. Mohamed.	Expert Systems with Applications.	2021	The paper offers a detailed exploration of extractive automatic text summarization (ATS) systems, addressing challenges and future research directions. It emphasizes the gap between computer-generated and human-generated summaries, highlighting the need for further advancements in ATS technology.
2.	Lucija Brezoˇcnik, Iztok Fister, Jr. and Vili Podgorelec	Applied Sciences	2018	The paper provides a detailed survey of 64 Swarm Intelligence (SI) algorithms for Feature Selection (FS), aiming to guide researchers and analysts in developing effective SI approaches for FS problems, filling a gap in the existing literature.
3.	Sourabh Katoch, Sumit Singh Chauhan & Vijay Kumar	Multimedia Tools and Applications	2020	The paper aims to review recent developments in genetic algorithms, highlighting key algorithms, genetic operators, and their applications. It focuses on facilitating new researchers and providing insights for their use in various domains.
4.	Ishitva Awasthi; Kuntal Gupta; Prabjot Singh Bhogal; Sahejpreet Singh Anand; Piyush Kumar Soni	6th International Conference on Inventive Computation Technologies (ICICT)	2021	The paper aims to review research on extractive and abstractive text summarization methods, highlighting challenges such as evaluation, labeled data, anaphora, and cataphora problems and proposing future directions for improvement.
5.	Laith Abualigah, Mohammad Qassem Bashabsheh, Hamzeh Alabool and Mohammad Shehab	Recent Advances in NLP: the case of Arabic language	2020	The paper aims to provide a comprehensive review of recent research in text summarization, focusing on abstractive summarization of multi-documents, and to outline the significance of text summarization in managing the vast amount of data on the web. Additionally, it aims to identify challenges in text summarization and propose future research directions in this field.
6.	Seyedali Mirjalili, Jin Song Dong, Ali Safa Sadiq and Hossam Faris	Nature Inspired Optimizers	2020	The paper aims to elucidate the core concepts of Genetic Algorithms (GA) and their application in different fields.
7.	Annu Lambora, Kunal Gupta, Kriti Chopra	International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)	2019	The paper aims to explain Genetic Algorithms (GA), including their basic workflow, functions, features, and applications. It also discusses theoretical concepts, advantages, and challenges, providing a comprehensive overview for readers.
8.	Roman Egger and Enes Gokce	Applied Data Science in Tourism	2022	The paper aims to highlight the importance of text data in academia and the tourism industry, discussing its challenges and opportunities. It also explores the historical development of text analysis and the role of Natural Language Processing (NLP) in data analysis and business operations.

9.	Kaiz Merchant; Yash Pande	2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)	2018	The paper aims to propose an automated text summarization system for legal judgments using latent semantic analysis (LSA) to improve accessibility to legal information and reduce the need for manual summarization.
10.	Lourdes Araujo	Genetic Programming and Evolvable Machines	2020	This paper explores the applications of genetic programming (GP) in natural language processing (NLP), reviewing existing works and identifying potential areas for further research. The paper also discusses the complementarity between GP and NLP, highlighting the need for more exploration in this field.
11.	Mohamed Atef Mosa, Arshad Syed Anwar, Alaa Hamouda	Knowledge-Based Systems	2019	This article aims to explore the application of swarm intelligence (SI) optimization techniques, particularly Ant Colony Optimization (ACO), in automatic text summarization, highlighting the potential and challenges in using SI for this task, especially in short text summarization (STS).
12.	Bilal, Millie Pant a, Hira Zaheer, Laura Garcia-Hernandez, Ajith Abraham	Engineering Applications of Artificial Intelligence	2020	This article aims to survey 25 years of Differential Evolution (DE), covering its basic aspects, customized variants, successful applications, and key bibliometric indicators. It seeks to generate interest among new users and guide experienced researchers in the optimization field.

Table 3 shows the 12 articles that have been selected for the understanding of the study.

Discussion

Swarm-Based Methods for Text Summarisation

Ant Colony Optimization (ACO)

The foraging action and behavior of ants lead to the concept of an efficient metaheuristic algorithm known as Ant Colony Optimization (ACO). Marco Dorigo first developed ACO at the beginning of the 1990s, and it is mostly utilized to resolve combinatorial optimization problems (Dorigo & Stützle, 2019). The Algorithm of ACO works in such a way that relies on the behavior of ants and their searching capacity for the shortest path from their nest to the desired food source, hence a process aided by forming and tracking pheromone trails (Neroni, 2021).

In ACO, a population of artificial ants continuously solves the optimization issue. Based on heuristic understanding and pheromone trails, each ant probabilistically chooses the subsequent part of the solution (Soofastaei, 2022). Ants develop solutions, and the quality of those solutions determines how often they update pheromone trails. Improved solutions are more likely to be adopted over time, resulting in the identification of superior solutions. An important characteristic of ACO is its capacity to manage complex and extensive optimization problems. Using the ant colony's collective intelligence, ACO can effectively search the space and identify almost ideal solutions. Numerous issues, such as task scheduling, vehicle routing, and the traveling salesman problem, have been effectively addressed by ACO (Lu & Yue, 2019).

The traveling salesman problem (TSP), in which a traveling salesman must visit each place just once, is frequently solved using the Ant colony optimization (ACO) metaheuristic. The goal is to identify the shortest possible Hamiltonian tour. Since the transfers between sites become components of the solution, the ACO metaheuristic is simple to understand (Wang & Han, 2021). Assigning the set of locations to the graph vertices defines the construction graph (GC). The ants build solutions by following the graph's edges, remembering their route, and selecting edges that don't go to vertices they have already visited. On the TSP, ant colony optimization has been demonstrated to work successfully. In addition, the set of vertices V, instead of the collection of edges E of the construction graph GC, can be associated with the set of solution elements of the TSP (or any other COP). It is possible to define the construction graph using any of the two proper concepts (Dorigo & Socha, 2018).

ACO's efficacy relies on its capacity to integrate exploration with exploitation. Ants start by exploring the search space and leaving pheromone trails. As the algorithm develops, pheromone trails become more intense on shorter pathways, directing additional ants to exploit them. This mix of exploration and exploitation allows ACO to effectively explore complicated optimization landscapes and discover high-quality solutions (Gao, 2020).

The flowchart of the ACO algorithm is presented in a rather comprehensive manner in the following figure:

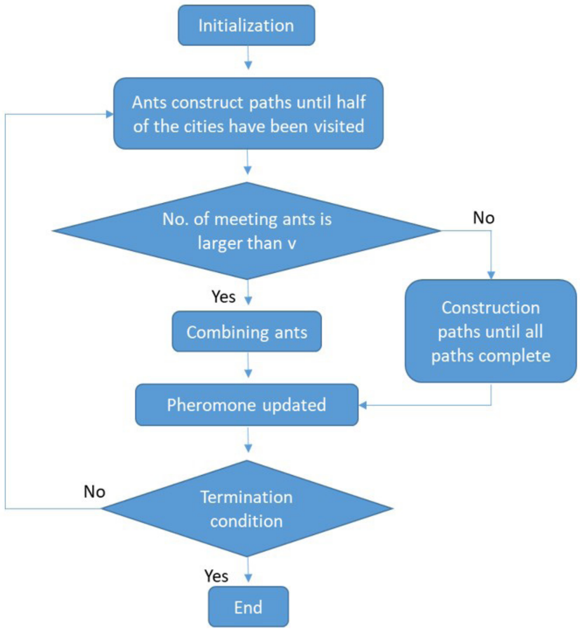

Figure 1: The flowchart of the Ant Colony Optimization (ACO) algorithm
Source: Author

The artificial ants develop solutions and update pheromone tracks to guide future investigation, which are essential parts of the ACO algorithm.
The calculation employed in ACO for updating the pheromone trails generally depends on the volume of pheromone produced by the ants and the pheromone's evaporation over time. A typical formula for updating pheromone trails is:

$$\tau ij=(1-\varrho)\ .\tau ij+\textstyle\sum_k \Delta\ \tau_{ij}^{k} \quad (1)$$

Where,

- τij is the pheromone concentration on the edge.
- ϱ is the rate at which pheromones evaporate.
- $\sum_k \Delta\ \tau_{ij}^{k}$ is the quantity of pheromone that ant k deposited on edge (i,j).

This equation depicts how pheromone trails are modified due to pheromone evaporation and pheromone deposits made by ants during solution building.

Particle Swarm Optimization (PSO)

Particle Swarm Optimization, discovered by Kennedy and Eberhart in 1995, is a versatile and effective metaheuristic algorithm applied due to its ease of use and performance in providing solutions to numerous optimization issues. The social behavior and actions of bird flocks and fish schools influenced the concept of PSO. It utilizes a population of particles that travel by a multidimensional searching space, changing their place based on their experiences and the swarms (Shami *et al.*, 2022). The main benefit of PSO is its easy deployment, which needs very few parameters to be set, making it easy and simple to adopt for numerous optimization tasks. Furthermore, the ability of PSO to access both local and global searches concurrently makes it more efficient in searching and meeting high-quality solutions (Wang *et al.*, 2018).

Although PSO has advantages, some disadvantages must be discussed to gain a better perspective of this algorithm. One significant disadvantage of PSO is that it tends to converge prematurely to inferior solutions, especially in high-dimensional and complicated optimization problems. This disadvantage is based on the algorithm's dependency on the best historical positions of particles, which will slow down the search process (Gad, 2022). Furthermore, the standard of solution provided by PSO might be sensitive to parameter selection, necessitating precise alteration for optimal performance.
A difference can be observed when comparing PSO and ACO, as they are derived from different social behavior concepts. ACO is inspired by the ant's behavior, whereas the behavior of fishes and flocks of birds inspires PSO. PSO is more frequently utilized for continuous optimization applications, but ACO works better for combinatorial optimization issues (Yarat *et al.*, 2021). Whereas PSO thrives in issues with continuous solution spaces, ACO often performs better when the search space is discontinuous, and the responses may be described as routes or sequences. Furthermore, because ACO builds and maintains pheromone trails, it usually takes more processing resources; in contrast, PSO is theoretically more efficient and better suited for large-scale optimization issues (Yarat *et al.*, 2021).
The flowchart of the PSO algorithm is presented in a rather comprehensive manner in the following figure:

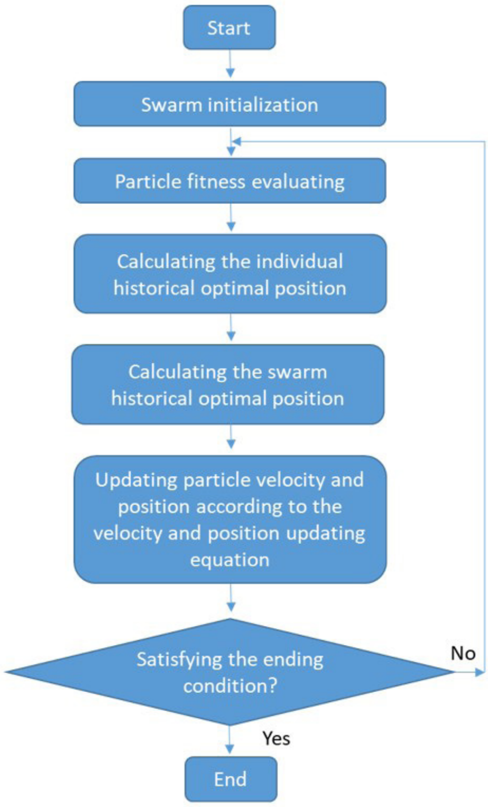

Figure 2: The flowchart of the Particle Swarm Optimization (PSO) algorithm
Source: Author

From a mathematical perspective, the PSO in the continuous space coordinate structure may be explained as follows:

$$vi,t+1=wvi,t+c1r1(x_{i,pbest,t}-xi,t)+c2r2(x_{g,best,t}-x_{i,t}) \quad (2)$$

Where,

- vi,t+1 shows the velocity of the particle (updated) i at time t+1.
- w represents the weight of inertia.
- The acceleration constants c1and c2show the cognitive and social components.
- r1 and r2 are the values taken randomly between 0 to 1,
- xi, best,t represents particles' personal best position i at time t,
- xg, best,t represents the swarm's global best position at time t,
- xi,t is the particle's current position i at time t.

Genetic-Based Methods for Text Summarisation

Genetic Algorithm (GA)

First proposed by Holland and his students in 1970, the genetic algorithm (GA) is based on the concept of evolution that mimics the natural selection procedure to search for an optimal solution to a given problem (Shen *et al.*, 2020). Genetic algorithms can be an important tool in text summarization to provide informative and concise summaries of the given information. The Algorithm of GA initiates with a population of the candidate's solutions, which are depicted as chromosomes usually encoded as binary strings.

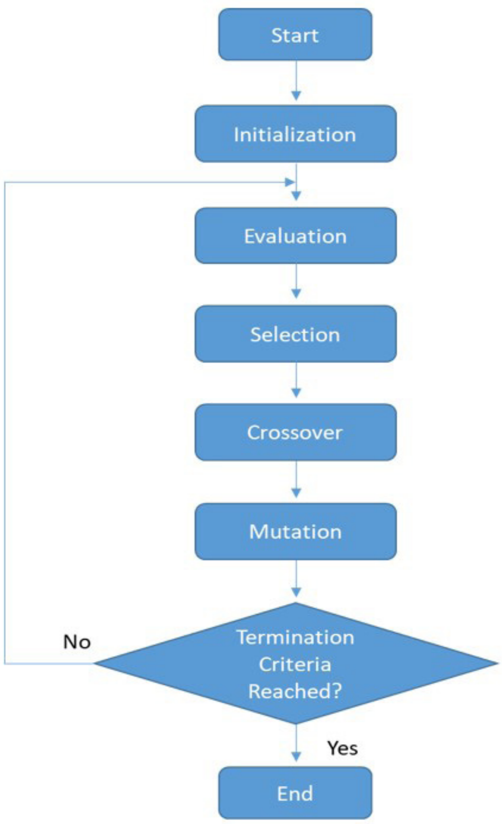

Figure 3: The flowchart of Genetic Algorithm (GA)
Source: Author

The depicted chromosomes undergo a series of genetic actions such as selection, crossover, and mutation to generate new offspring (Katoch *et al.*, 2021).

The offspring generated from the genetic operations is assessed using a fitness function that measures the summary by capturing the significant knowledge from the original text. The individuals that are the fittest from the offspring are selected for the formation of the next generation, which repeats continuously until the termination phase occurs so that it reaches a limited number of generations or attains a satisfactory summary (Lambora *et al.*, 2019). The GA is a robust and adaptive text summarizing technique that can solve complicated issues in huge search areas. However, its performance is highly dependent on choosing parameters and fitness function approach, which might result in weak solutions or slower convergence. Furthermore, GA can be computationally costly, particularly with big datasets or sophisticated summarizing tasks.

Genetic Programming (GP)

Developed in 1962 by Koza, summarizing text using Genetic Programming (GP) is a technique that employs evolutionary algorithms, frequently developed from genetic algorithms (GA), to construct short and helpful overviews of text materials. In this method, a computer program describing the summary (phenotype) is recorded as a genome (genotype), which can be a syntax tree, a command sequence, or other hierarchical structures for data (Ahvanooey *et al.*, 2019). The fitness function assesses the value of the summary generated by the program using pre-set criteria such as relevancy, coherence, and length.

The application of GP for text summarizing is helpful since it automates the summary process, conserving time and effort over manual techniques. It is also adjustable to varied types of text and summary needs, making it suitable for a wide range of applications. GP can be scaled to accommodate vast amounts of text, making it ideal for summarizing massive publications or data sets. Furthermore, GP provides flexibility in creating summaries in various forms and styles, enabling customization based on individual summarizing requirements.

However, there are certain downsides to employing GP for text summarizing. Processing times may be prolonged by GP's computational complexity, particularly when dealing with big documents or datasets. When an evolved program works well on training data but fails to translate well to new, unknown data, it will likely be overfitting. Biases may be introduced into the produced summaries due to the subjective character of the fitness function used to assess summaries (Bajaj & Sangwan, 2019). Moreover, complicated and challenging-to-understand evolving programs in GP may impede comprehension of the summaries' generation process. Hence, GP necessitates substantial computing resources, such as memory and processing capacity, which can restrict its use in some scenarios (Fu *et al.*, 2023).

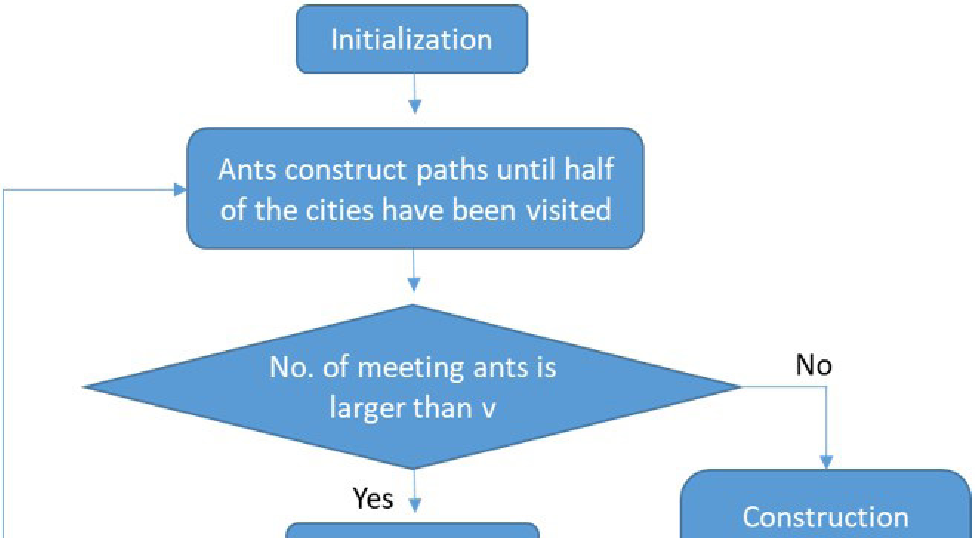

Figure 4: The flowchart of Genetic Programming (GP)
Source: Author

Differential Evolution (DE)
Differential Evolution (DE) is an approach based on population optimization that can be applied to text summarization (Meng *et al.*, 2019). DE is influenced by natural selection and the survival of the fittest. It develops a population of potential solutions across generations, depicting each solution as a vector in a highly dimensional search space. DE can be used in text summarizing to improve the selection of phrases or sentences that best reflect a document's core content. The method begins with a starting point of candidate summaries, generally represented as a collection of words or sentences from the document. These potential summaries are then assessed using a fitness function that determines how well it summarize the material (Lynn *et al.*, 2018).

The simplicity of DE is one of its primary advantages. Unlike other optimization algorithms, it is simple to develop and requires less adjusting because it has fewer algorithmic parameters. Compared to conventional optimization techniques, DE usually takes fewer function executions to converge to a solution, making it computationally efficient. This efficiency is particularly useful in high-dimensional optimization issues where it is not practicable to use exhaustive search methods (Deng *et al.*, 2021). Another benefit of DE is its ability to withstand noise and nonlinearities in its intended function (Ahmad *et al.*, 2022). DE's population-based strategy enables it to explore various parts of the search space simultaneously, preventing it from being stuck in the optimal area. Furthermore, DE does not require the objective function's gradient information, making it appropriate for optimizing black-box variables with an unknown or difficult-to-calculate analytical form (Yuen *et al.*, 2019).

Despite its benefits, DE does have some drawbacks. One disadvantage of DE is that its performance can be affected by parameter options such as population size, mutation method, and crossover probability (Caraffini *et al.*, 2019). Poorly chosen parameters might result in premature or weak convergence to a solution. Furthermore, DE can have trouble with multimodal optimization issues in which the objective function has many local optima. It may not adequately traverse the search space to identify the global optimum (Chen *et al.*, 2019).

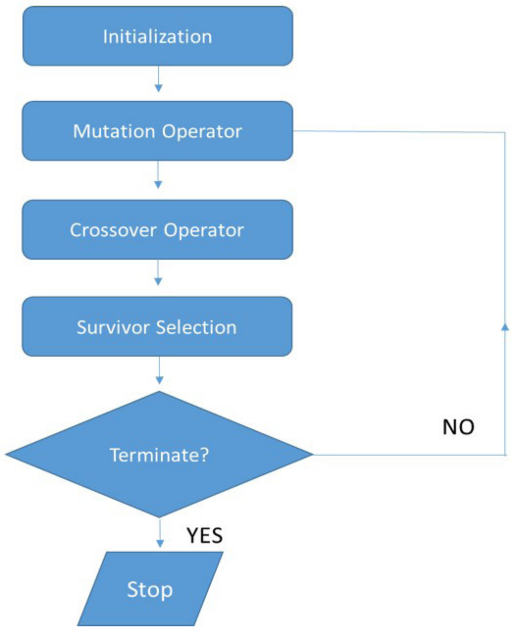

Figure 5: The flowchart of Differential Evolution (DE)
Source: Author

Performance Comparison of Swarm-Based and Genetic-Based Methods
Genetic-based and swarm-based techniques are widely used for text summarization, and each has advantages and disadvantages. Influenced by natural swarm behavior, swarm-based techniques like Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) employ a set of basic principles to help individual agents find optimum

solutions collectively (Gad, 2022). To evolve solutions across generations, genetic-based techniques such as Genetic Algorithms (GA) imitate the natural selection procedure (Rostami, Berahmand, & Forouzandeh, 2021). Concerning performance, both techniques have demonstrated promise in text summarizing tasks, but they have distinct properties. Swarm-based approaches are frequently praised for their rapid convergence to a solution and resilience to local optima. They are very useful for continuous optimization issues and can effectively deal with noisy objective functions (Bansal *et al.*, 2019). However, they might encounter complicated search spaces or discrete-solution issues, typical in text summarization. On the other hand, genetic-based approaches are well-known for their capacity to test various solutions while maintaining population diversity. This can be useful for text summarization since it enables a more thorough investigation of the search space. Genetic algorithms are also adaptable to diverse problem areas by varying the genetic operators and variables (Rostami, Berahmand, & Forouzandeh, 2021). However, they may take more computer resources and effort to converge than swarm-based approaches.

The particular needs and limitations of the application will determine whether to use genetic or swarm-based techniques for text summarization. Swarm-based methods, like PSO and ACO, are commended for being simple and easy to use, which makes them appropriate for applications with constrained computational resources (Chen & Shang, 2021). Additionally, because of their natural parallelism, they may be readily parallelized to take advantage of distributed or multi-core computing systems. GA and other genetic-based methods are more versatile and flexible, making them appropriate for complex optimization problems or circumstances where the search space is not widely understood (Gen & Lin, 2023b). Even so, this flexibility occurs at the expense of more complexity and the necessity for more precise parameter modification. To summarize, the decision between swarm-based and genetic-based text summarizing approaches is determined by the nature of the issue, available computational capacity, and the ideal combination of simplicity and flexibility.

Challenges and Problems of Using Swarm-Based and Genetic-Based Methods in Text Summarisation

Table 4: Challenges and problems of using Swarm-Based and Genetic-Based Methods in text summarization

S. no	Challenges/Problems	Swarm-Based Methods	Genetic-Based Methods
1.	Scalability and Computational Complexity	Larger datasets demand more computational resources, which could lead to scaling issues (Dehghani *et al.*, 2021)	In comparison to swarm-based techniques, genetic algorithms could take longer to converge and demand additional computational resources (Rahmani Hosseinabadi *et al.*, 2019).
2.	Parameter Initialisation and Settings	Characteristics like population size and mutation rate must be carefully selected and tuned to achieve optimal performance.	Genetic algorithms need adjustment of genetic operators and parameters, which increases complexity (Gen & Lin, 2023a).
3.	Partial Adaptability for Diverse Data	Difficulty adjusting to various text data, such as languages, styles, or themes.	Genetic algorithms may fail to adapt to different data sets and problem contexts.
4.	Interpretability of Generated Text-Based Summaries	Users may find it difficult to read summaries that have been generated because of their lack of interpretability.	Because evolution is a complicated process, it can be difficult to interpret the results of genetic algorithms (Gen & Lin, 2023a).

Limitations

• One of the review's limitations is a lack of data focusing exclusively on Genetic-Based approaches in NLP text summarization instead of Swarm-Based approaches. This discrepancy might make it more difficult to compare the two strategies in depth.

• Furthermore, the speed at which technology develops might eventually affect the findings' application and significance.

Future Implications

• Future work in Swarm-Based and Genetic-Based text summarizing techniques should concentrate on several important issues.

• For more accurate evaluations and result comparisons, standardized evaluation measures are required to evaluate the effectiveness of various summarizing techniques reliably.

• Investigating hybrid strategies that include these techniques with deep learning or machine learning may improve text summarization's efficacy and efficiency.

CONCLUSION

Finally, the comparative study of Swarm-Based and Genetic-Based approaches for text summarization in Natural Language Processing (NLP) provides useful insights. Swarm-based approaches, such as particle swarm optimization and ant colony optimization, provide efficient search tactics, whereas genetic-based methods, such as genetic algorithms, use natural selection processes to provide unique solutions. Despite the distinctions, both techniques offer benefits and drawbacks that affect their applicability for particular summarization tasks. Future research should focus on standardizing assessment measures, investigating hybrid techniques, and applying these methods to real-world applications to improve utility and performance. Overall, this paper advances NLP by thoroughly comparing different approaches and indicating opportunities for future research and growth in text summarizing.

Acknowledgment

The author thanks the University of Hafr Al Batin for their support and resources in this research endeavor.

REFERENCES

Abualigah, L., Bashabsheh, M. Q., Alabool, H., & Shehab, M. (2020). Text summarization: A brief review. *Recent Advances in NLP: The case of Arabic language,* 1–15.

Abuobieda, A., Salim, N., Albaham, A. T., Osman, A. H., & Kumar, Y. J. (2012). Text summarization features selection method using pseudo genetic-based model. In *2012 International Conference on Information Retrieval & Knowledge Management.*

Abuobieda, A., Salim, N., Binwahlan, M. S., & Osman, A. H. (2013). Differential evolution cluster-based text summarization methods. In *2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE).*

Abuobieda, A., Salim, N., Eltayeb, R., Bin Wahlan, M. S., Suanmali, L., & Hamza, A. (2011). Pseudo genetic and probabilistic-based feature selection method for extractive single document summarization. *Journal of Theoretical and Applied Information Technology, 32*(1), 80–87.

Abuobieda, A., Salim, N., Kumar, Y. J., & Osman, A. H. (2013). An improved evolutionary algorithm for extractive text summarization. In *Intelligent Information and Database Systems: 5th Asian Conference, ACIIDS 2013, Kuala Lumpur, Malaysia, March 18–20, 2013, Proceedings, Part II,* 5.

Ahmad, M. F., Isa, N. A. M., Lim, W. H., & Ang, K. M. (2022). Differential evolution: A recent review based on state-of-the-art works. *Alexandria Engineering Journal, 61*(5), 3831–3872.

Ahvanooey, M. T., Li, Q., Wu, M., & Wang, S. (2019). A survey of genetic programming and its applications. *KSII Transactions on Internet and Information Systems, 13*(4), 1765–1794.

Awasthi, I., Gupta, K., Bhogal, P. S., Anand, S. S., & Soni, P. K. (2021). Natural language processing (NLP) based text summarization: A survey. In *2021 6th International Conference on Inventive Computation Technologies (ICICT).*

Bajaj, A., & Sangwan, O. P. (2019). A systematic literature review of test case prioritization using genetic algorithms. *IEEE Access, 7,* 126355–126375.

Bansal, J. C., Singh, P. K., & Pal, N. R. (2019). *Evolutionary and swarm intelligence algorithms* (Vol. 779). Springer.

asha, M. J., Vijayakumar, S., Jayashankari, J., Alawadi, A. H., & Durdona, P. (2023). Advancements in natural language processing for text understanding. *E3S Web of Conferences, 321,* 01056.

Boorugu, R., & Ramesh, G. (2020). A survey on NLP-based text summarization for summarizing product reviews. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 858–863). IEEE.

Brezočnik, L., Fister Jr, I., & Podgorelec, V. (2018). Swarm intelligence algorithms for feature selection: A review. *Applied Sciences, 8*(9), 1521.

Caraffini, F., Kononova, A. V., & Corne, D. (2019). Infeasibility and structural bias in differential evolution. *Information Sciences, 496,* 161-179.

Chen, Y., & Shang, N. (2021). Comparison of GA, ACO algorithm, and PSO algorithm for path optimization on free-form surfaces using coordinate measuring machines. *Engineering Research Express, 3*(4), 045039.

Chen, Z.-G., Zhan, Z.-H., Wang, H., & Zhang, J. (2019). Distributed individuals for multiple peaks: A novel differential evolution for multimodal optimization problems. *IEEE Transactions on Evolutionary Computation, 24*(4), 708–719.

Dehghani, M., Hubálovský, Š., & Trojovský, P. (2021). Northern goshawk optimization: a new swarm-based algorithm for solving optimization problems. *IEEE Access, 9,* 162059-162080.

Deng, W., Shang, S., Cai, X., Zhao, H., Song, Y., & Xu, J. (2021). An improved differential evolution algorithm and its application in optimization problem. *Soft Computing, 25,* 5277-5298.

Dorigo, M., & Socha, K. (2018). An introduction to ant colony optimization. In *Handbook of approximation algorithms and metaheuristics* (pp. 395-408). Chapman and Hall/CRC.

Dorigo, M., & Stützle, T. (2019). *Ant colony optimization: Overview and recent advances.* Springer.

Egger, R., & Gokce, E. (2022). Natural Language Processing (NLP): An Introduction: Making Sense of Textual Data. In *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 307-334). Springer.

El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications, 165,* 113679.

Fu, X., Sun, Y., Wang, H., & Li, H. (2023). Task scheduling of cloud computing based on hybrid particle swarm algorithm and genetic algorithm. *Cluster Computing, 26*(5), 2479-2488.

Gad, A. G. (2022). Particle swarm optimization algorithm and its applications: A systematic review. *Archives of Computational Methods in Engineering, 29*(5), 2531–2561.

Gao, W. (2020). New ant colony optimization algorithm for the traveling salesman problem. *International Journal of Computational Intelligence Systems, 13*(1), 44-55.

Gen, M., & Lin, L. (2023a). Genetic algorithms and their applications. In *Springer handbook of engineering statistics* (pp. 635-674). Springer.

Gen, M., & Lin, L. (2023b). Nature-inspired and evolutionary techniques for automation. In *Springer Handbook of Automation* (pp. 483-508). Springer.

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications, 80,* 8091–8126.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends, and challenges. *Multimedia Tools and Applications, 82*(3), 3713–3744.

Kieuvongngam, V., Tan, B., & Niu, Y. (2020). Automatic text summarization of COVID-19 medical research articles using BERT and GPT-2. *arXiv preprint arXiv:2006.01997.*

Lambora, A., Gupta, K., & Chopra, K. (2019). Genetic algorithm—A literature review. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 162–168). IEEE.

Lu, L.-C., & Yue, T.-W. (2019). Mission-oriented ant-team ACO for min–max MTSP. *Applied Soft Computing, 76,* 436–444.

Lynn, N., Ali, M. Z., & Suganthan, P. N. (2018). Population topologies for particle swarm optimization and differential evolution. *Swarm and Evolutionary Computation, 39,* 24–35.

Manegre, M., & Sabiri, K. A. (2022). Online language learning using virtual classrooms: An analysis of teacher perceptions. *Computer Assisted Language Learning, 35*(5-6), 973-988.

Meng, Z., Pan, J.-S., & Tseng, K.-K. (2019). PaDE: An enhanced Differential Evolution algorithm with novel control parameter adaptation schemes for numerical optimization. *Knowledge-Based Systems, 168,* 80-99.

Merchant, K., & Pande, Y. (2018). NLP-based latent semantic analysis for legal text summarization. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1369–1374). IEEE.

Mosa, M. A. (2020). A novel hybrid particle swarm optimization and gravitational search algorithm for multi-objective optimization of text mining. *Applied Soft Computing, 90,* 106189.

Nazari, N., & Mahdavi, M. (2019). A survey on automatic text summarization. *Journal of AI and Data Mining, 7*(1), 121-135.

Neroni, M. (2021). Ant colony optimization with warm-up. *Algorithms, 14*(10), 295.

Rahmani Hosseinabadi, A. A., Vahidi, J., Saemi, B., Sangaiah, A. K., & Elhoseny, M. (2019). Extended genetic algorithm for solving open-shop scheduling problem. *Soft Computing, 23,* 5099-5116.

Rane, N. (2023). Role and challenges of ChatGPT and similar generative artificial intelligence in business management. *SSRN.*

Rostami, M., Berahmand, K., & Forouzandeh, S. (2021). A novel community detection-based genetic algorithm for feature selection. *Journal of Big Data, 8*(1), 1–27.

Rostami, M., Berahmand, K., Nasiri, E., & Forouzandeh, S. (2021). Review of swarm intelligence-based feature selection methods. *Engineering Applications of Artificial Intelligence, 100,* 104210.

Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle swarm optimization: A comprehensive survey. *IEEE Access, 10,* 10031-10061.

Shen, X., Zheng, Y., & Zhang, R. (2020). A hybrid forecasting model for the velocity of hybrid robotic fish based on back-propagation neural network with genetic algorithm optimization. *IEEE Access, 8,* 111731-111741.

Sohail, A. (2023). Genetic algorithms in the fields of artificial intelligence and data sciences. *Annals of Data Science, 10*(4), 1007-1018.

Soofastaei, A. (2022). Introductory chapter: Ant colony optimization. In *The application of ant colony optimization* (pp. 1–10). IntechOpen.

Tsamardinos, I., Borboudakis, G., Katsogridakis, P., Pratikakis, P., & Christophides, V. (2019). A greedy feature selection algorithm for Big Data of high dimensionality. *Machine learning, 108,* 149-202.

Vashisht, V., Pandey, A. K., & Yadav, S. P. (2021). Speech recognition using machine learning. *IEIE Transactions on Smart Processing & Computing, 10*(3), 233-239.

Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: An overview. *Soft Computing, 22,* 387–408.

Wang, Y., & Han, Z. (2021). Ant colony optimization for traveling salesman problem based on parameters optimization. *Applied Soft Computing, 107,* 107439.

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., & Affandy, A. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University-Computer and Information Sciences, 34*(4), 1029–1046.

Yang, B., Luo, X., Sun, K., & Luo, M. Y. (2023). Recent progress on text summarisation based on BERT and GPT. In *International Conference on Knowledge Science, Engineering and Management* (pp. 1–9). Springer.

Yarat, S., Senan, S., & Orman, Z. (2021). A comparative study on PSO with other metaheuristic methods. In *Applying Particle Swarm Optimization: New Solutions and Cases for Optimized Portfolios* (pp. 49–72). Springer.

Yuen, S. Y., Lou, Y., & Zhang, X. (2019). Selecting evolutionary algorithms for black-box design optimization problems. *Soft Computing, 23*(15), 6511–6531