Jordanian Journal of Informatics and Computing

Unsupervised text feature selection approach based on improved Prairie dog algorithm for the text clustering

Mohammad Alshinwan[1], Abdul Ghafoor Memon[2], Mohamed Chahine Ghanem[3], Mohammed Almaayah[4]

[1] *Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan.*
[2] *IMCS, University of Sindh, Jamshoro, Pakistan.*
[3] *Cyber Security Research Centre, London Metropolitan University, London, UK*
[4] *King Abdullah the II IT School, The University of Jordan, Amman 11942, Jordan*

ARTICLE INFO

Article History
Received: 01 March 2025
Revised: 22 March 2025
Accepted: 24 March 2025
Published: 25 March 2025

Academic Editor:
Shahed Almobydeen

Vol.2025, No.1

DOI:

ABSTRACT

Text clustering is suitable for dividing many text documents into distinct groups. The size of the documents has an impact on the performance of text clustering, reducing its effectiveness. Text documents often include sparse and uninformative characteristics, which can negatively impact the efficiency of the text clustering technique and increase the computational time required. Feature selection is a crucial strategy in unsupervised learning that involves choosing a subset of informative text features to enhance the efficiency of text clustering and decrease computing time. This work presents a novel approach based on an improved Prairie dog algorithm to solve the feature selection problem. K-means clustering is employed to assess the efficacy of the acquired subgroups of features. The proposed algorithm is being compared to other algorithms published in the literature. The feature selection strategy ultimately promotes the clustering algorithm to get precise clusters.

Keywords: Feature selection, Data clustering, Document clustering, Optimization, PDO algorithm.

How to cite the article

1. Introduction

Clustering is a popular unsupervised learning technique with many applications in many scientific fields, including text processing, pattern recognition, image processing, and information retrieval. Clustering involves assembling sets of data that are very similar to one another into more prominent groups to reduce the degree of similarity between those sets. There are two main types of existing clustering methods: hierarchical and partitional. When using hierarchical clustering, the final result will be a collection of nested clusters displayed as hierarchical trees. Each data item is guaranteed to be in a unique subset of the dataset when using partitional clustering.[1][2][3]. It has been applied in various fields within text mining, including text retrieval, text categorization, human movement analysis, fall detection, and image segmentation [4][5]. The vector space model (VSM) is widely used in text mining and represents the text properties of each document as a vector of term weights. Each term weight in this model is represented as a one-dimensional space [6][7].

*Corresponding author. Email: m_shinwan@asu.edu.jo

Text clustering is a highly effective method of unsupervised learning used to handle large numbers of text documents without prior knowledge of their class labels [8] [9]. This method divides a collection of substantial textual documents into organized and meaningful groups by gathering related documents that share similar qualities inside the cluster, as determined by their inherent properties [10].The identical clusters encompass pertinent and analogous text documents. Conversely, several clusters have unrelated and distinct text documents [11].

Clustering is a crucial task today due to the vast amount of online textual data [12]. Clustering is employed to identify pertinent text documents and streamline their presentation by grouping those that exhibit similar patterns and content. The text clustering technique is effectively employed in various study domains to streamline the process of text analysis, including data mining, digital forensics investigation, and information retrieval [11].

The vector space model (VSM) is the prevailing model employed in text classification (TC) to depict each document. In this instance, each term (word) in the term-document matrix serves as a feature (word) for representing the document The TDs are depicted in a multi-dimensional space, where the position value of each measurement coordinates to a term frequency and commonness value [13][3]. A tiny document might yield hundreds or even thousands of text characteristics produced from various text phrases. Therefore, TDs will possess informative and uninformative features in high-dimensional space, including unevenly distributed, redundant, noisy, and irrelevant features. The feature selection (FS) technique can be used to delete these uninformative characteristics [14] [15].

Feature selection approaches are optimization methods that are nondeterministic polynomial time hard. They are used to identify the optimal subset of essential text features and enhance the performance of the TC method while still preserving the critical text data [16]. Usually, these methods are executed without prior knowledge of the document's class label. Traditionally, these strategies are categorized into three primary types: Feature selection based on document frequency, Feature selection based on term frequency, and a hybrid feature technique based on both term and document frequency. Various text-oriented studies utilize feature selection (FS) techniques, including topic clustering, text categorization, and knowledge discovery. Metaheuristic algorithms have recently proven effective in the field of text mining for solving text feature selection and text document clustering [17][18].

Meta-heuristic optimization techniques have been widely used for many optimization issues over the past two decades [19] Meta-heuristic algorithms have effectively been employed in the field of text mining to address text feature selection challenges. Text feature selection methods, such as prairie dog algorithm [20], arithmetic optimization algorithm [21], Ant lion optimizer [22], and grey wolf optimizer [23], have been popular in research. These methods are population-based optimization algorithms. These algorithms strive to achieve enhanced solutions by leveraging knowledge gained from earlier iterations. Various meta-heuristic algorithms are employed for solving feature selection problems.

Proposing a new feature approach and improving the performance of feature selection techniques to get satisfying results are the primary goals of this paper. Specifically, we aim to achieve the following:

- Identifying the most useful features per document requires a new approach, and one way to do it is to use an optimization strategy. With this technique, we can speed up text clustering and cut down on the amount of time it takes to process data.

- To find the best relevant and comparable document clusters using the acquired redaction subset in the K-mean clustering technique.

The remainder of this paper is structured as follows: Section 2 discusses the relevant research and the reasons for proposing this study. In Section 3 of this study, the proposed work is illustrated. The findings are detailed in Section 4. Finally, section 5 concludes this study.

2. Related Works

2.1 Previous studies

Abualigah et al.[24] offered a wide range of textures and a vivid degradation of dimensions for text grouping. Text clustering techniques are widely employed for classifying textual data. Text clusters are utilized in several applications, such as text mining, information retrieval, and classification. Text feature sampling is crucial for the process of text clustering. The text information possesses both informative and non-informative capabilities. These misleading traits deceive and reduce their dependability for classification algorithms and techniques. The utilization of the harmony search algorithm, swarm particle optimization, and genetic algorithm has selected the text features. The function selection procedures involved the incorporation of a weighting scheme known as length-weight. The dynamic dimensional reduction technique decreased the number of characteristics utilized for clustering. The text document is subjected to clustering using the k-means clustering algorithm. The proposed method is enhanced compared to the current technology.

In this work [25], the authors utilize the Grasshopper Optimization Algorithm (GOA) as a search technique for designing a feature selection method based on wrappers. The Grasshopper Optimization Algorithm (GOA) is a novel metaheuristic approach that emulates the collective movement patterns of grasshopper populations. This study introduces a very effective optimizer that combines Evolutionary Population Dynamics (EPD), selection operators, and Genetic Optimization Algorithm (GOA). The optimizer is presented in four distinct methods to address the limitations of the conventional GOA, specifically its tendency to converge prematurely and become stagnant.

Aljarah et al. suggest a hybrid methodology utilizing the grasshopper optimization algorithm (GOA), a novel method that mimics the collective behavior observed in swarms of grasshoppers. The objective of the suggested method is to enhance the parameters of the SVM algorithm and identify the most optimal subset of features concurrently. The suggested approach is evaluated for accuracy using eighteen benchmark data sets, which vary in low and high dimensions. The proposed technique is compared with seven reputable algorithms. In addition, the suggested method is compared to grid search, the prevailing strategy for optimizing SVM parameters [26].

This study [27] presents an improved binary grey wolf optimizer (GWO) version using a wrapper feature selection (FS) technique to address Arabic text categorization challenges. The binary GWO algorithm is employed as a wrapper-based feature selection strategy. An investigation is conducted to evaluate the performance of the suggested method utilizing various learning models, such as decision trees, K-nearest neighbor, Naive Bayes, and SVM classifiers. The effectiveness of various BGWO-based wrapper approaches is assessed using three Arabic public datasets: Alwatan, Akhbar-Alkhaleej, and Al-jazeera-News.

2.2 Motivation of this work:

The meta-heuristic algorithms draw inspiration from animal behavior and natural phenomena. Meta-heuristic algorithms operate at the population level and possess the capacity to mitigate the stagnation of local optima to a certain degree. Additionally, it has strong convergence capabilities towards optimal solutions. Overall, meta-heuristic algorithms demonstrate a significant inclination toward exploitation. Nevertheless, it may not consistently execute exploration effectively. Therefore, in certain instances, meta-heuristic algorithms may not be able to effectively address the problem and are unable to identify the global optimal solution. The prairie dog algorithm (PDO) is a newly suggested meta-heuristic approach that imitates the intelligent actions of groups of prairie dogs [28]. The local search approach employed by the PDO is highly effective. This study introduces a PDO algorithm to strike a better balance between diversification and intensification, resulting in improved solution accuracy and convergence speed compared to other meta-heuristic algorithms.

3. Proposed work:

In text clustering, a collection of large text documents is transformed into a smaller set of lattices based on their underlying properties. Prior to employing clustering algorithms, a benchmark preparation step is employed to preprocess text content. To reduce computing time and prioritize enhancing text clustering defense, text pretreatment techniques are utilized to identify the most effective thought-stimulating elements. The proposed method identifies a novel subset of informationoriented text features characterized by limited spaces. The primary benefit of the PDO algorithm, in comparison to other widely recognized meta-heuristic algorithms, is its ability to operate without the need for any specified input parameters. Furthermore, it is simple and devoid of computational intricacies. Additionally, its benefits encompass the simplicity of converting this programming language concept and the ease of understanding it. PDO excels at effortlessly circumventing local optima in the face of intricate, high-dimensional, and multimodal issues. Figure 1 depicts the structure of the proposed technique.

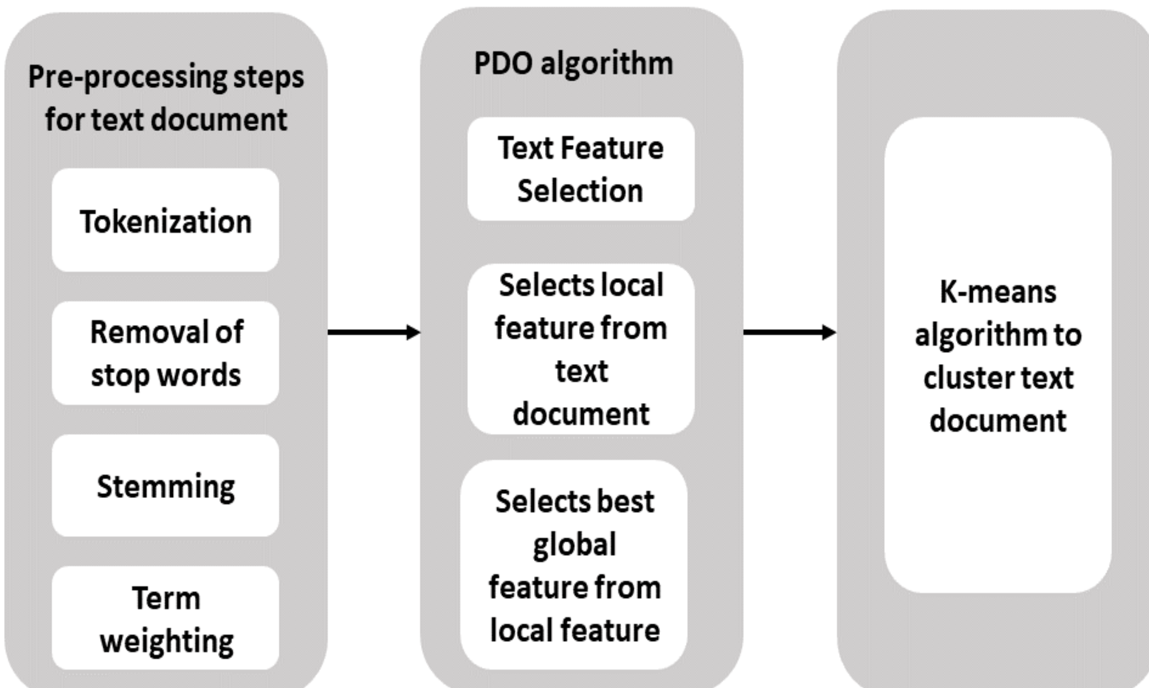

Figure 1. The structure of the proposed technique.

3.1 Pre-processing of text documents

To effectively utilize various text mining techniques, such as clustering, feature selection, and retrieval of text, it is necessary to convert the contents of documents into a format that can be easily processed by the underlying algorithm [29]. The preprocessing stages are employed to transform the textual contents of the document into a numerical representation. The steps are categorized as follows:

A. 3.1.1. Tokenization

Tokenization involves the procedure of dividing a text flood into individual terms or phrases and eliminating any empty spaces. Each word, term, and symbol is selected sequentially from the beginning to the end of the text and is called a token.

B. 3.1.2. Stop words.

The stop words refer to a collection of often used and short functional words, such as "an," "this," "that," "when," and "be," which carry little significance in text document clustering and are therefore assigned low weights. These words should be eliminated from documents as they typically occupy a portion of the content and impact the growing number of features, thus resulting in diminished performance of the text clustering approach.

C. 3.1.3. Stemming

Stemming, by removing prefixes and suffixes, imbues the conjugated verbs of certain nouns with a nearly identical root. The terms effective, productive, and effectual all share the same origin: "effect," indicating a characteristic or quality.

D. 3.1.4. Term weighting

The procedure of term weighting involves converting data from its textual to numerical form. The term frequency-inverse document frequency (TFIDF) is used to estimate the term weighting for document representation in the field of text mining. [30].

3.1 PDO algorithm

In the last two decades, numerous advanced algorithms have been developed due to the continuous advancement of natural-inspired optimization methods. However, there is still a need to evaluate their effectiveness in tackling progressively complex optimization problems. There are two main categories of optimization methods: deterministic and non-deterministic (stochastic). Deterministic methods can be further classified into linear and nonlinear approaches. This study aims to evaluate the utility and effectiveness of the Prairie Dog Optimization algorithm (PDO), a newly developed optimization technique, in addressing engineering structural damage detection problems.

The behavior of prairie dogs elicits PDO behavior in their natural habitat. PDO utilizes four prairie dog behaviors to accomplish the two primary research and management optimization stages. Prairie dogs utilize their grazing and burrowing actions to facilitate the assessment of PDO performance. Prairie dogs construct burrows in close proximity to abundant food sources. In response to the depleting food supply, they actively seek an alternative source of provisions, construct fresh underground dwellings within it, and thoroughly investigate the entire community or designated area to locate a new source of rations or potential alternatives. This strategy capitalizes on the distinct reactions of prairie dogs to various communication or alarm sounds. Prairie dogs possess a range of vocalizations or signals with distinct functions, such as indicating the presence of predators or food availability. Prairie dogs possess practical communication skills to fulfill their nutritional needs and protect themselves from predators. Within the framework of PDO implementation, these discrete events stimulate the prairie dogs to congregate at particular locations or potential sites. Further search operations, referred to as exploitation, are carried out to discover enhanced or nearly optimum solutions [28].

Prairie dogs (PD) are acknowledged by the scientific community for having one of the most advanced and all-encompassing communication systems among animals. Although PD vocalizations may appear uncomplicated to humans, they effectively communicate complex signals to other PDs. These creatures possess sophisticated communication skills, enabling them to assign specific names to individual predators. PDs often adopt an upright posture or raise their heads while feeding in order to remain vigilant for potential threats. Upon detecting a threat, these creatures make vocalizations resembling nasal yips and promptly withdraw into their burrows.

PDs demonstrate intricate and extraordinary anti-predation responses. Due to their dichromatic color vision, they can detect predators from a distance and notify nearby PDs of imminent danger. PDs generate distinctive high-pitched sounds that enable other species members to identify particular predators and gauge their approaching velocity. Moreover, the reactions of PDs differ based on the specific predator involved. When a hawk is sighted, neighboring PDs cease their activity and assume an upright position on their hind legs to survey the direction of the impending attack. Individuals in the route of the hawk swiftly withdraw inside their burrows, while others watch from the entrances of their burrows. Likewise, group members quickly move towards the burrow's entrance when coyotes are about and remain vigilant. The suggested PDO is based on the particular response of PDs to different predators.

PDO exploits the continuous pattern of actions exhibited by PDs, utilizing a distinct response to auditory stimuli. Various animals react accordingly to different sounds, analyzing them concerning signals that suggest possible positions. The recommended PDO is built upon this ongoing sequence of actions.

The different elements initiate the optimization process by moving from one food source to another in search of nourishment. PDs are predominantly herbivorous, although they may occasionally consume some insects. Throughout the year, they migrate between different locations, mostly feeding on grasses, small seeds, and specific insects while actively searching for issues [28]. Algorithm 1 summarizes the PDO algorithm pseudo code.

Algorithm 1: Pseudo-code for PDO Algorithm

Require: N, T, LB, UB, Dim, F_obj
Ensure: $Best\_PD, PDBest\_P, PDConv$

```
{Initialize all necessary variables}
{Evaluate fitness for each solution and record the best one}
while not exceeding maximum iterations do
    {Calculate DS, PE, RL, TPD according to the current phase of t}
    for each solution do
        {Calculate cpd, P, eCB}
        {Update solution depending on the current phase of t}
        {Apply boundary conditions to solution}
        {Calculate new fitness}
        if new fitness is better then
            {Update solution and its fitness}
        end if
        if current fitness is globally best then
            {Update global best fitness and solution}
        end if
    end for
    {Update convergence curve}
    {Display the best solution fitness every 50 iterations}
    {Increment iteration count}
end while
```

The PDO function represents an optimization algorithm inspired by the behavior of the dung beetle. The dung beetle represents one of the most fascinating examples of animal behavior, as they are known to navigate in a straight line, no matter what obstacles they encounter.

Parameters:

- N: The number of beetles (solutions).
- T: The maximum number of iterations.
- LB and UB: The lower and upper bounds of the search space.
- Dim: The dimensionality of the search space.
- F_obj: The objective function used to evaluate the fitness of a solution.
- DS: The Digging strength variable which is influenced by the iteration progress and a random number.
- PE: The Predator effect variable which is influenced by the iteration progress.
- RL: The Levy random number vector.
- TPD: The position of the best solution so far.

4. Experimental results.

4.1. Dataset:

Table 1 displays the eight different typical standard text datasets utilized to assess the execution of the suggested PDO approach. Numerical representations of text clustering standard datasets can be found in this link, following the extraction of words.

Table 1 The datasets used in the experimental results.

Datasets	Number of documents	Number of terms	Number of clusters
DS_1	2	2935	2
DS_2	1	3263	4
DS_3	1	263	7
DS_4	2	5773	9
DS_5	3	312	1
DS_6	1	265	12
DS_7	5	9641	15
DS_8	1	1935	2

4.2 Performance evaluation

In this section, we compare the proposed algorithm's performance to that of another comparative algorithm in the feature selection domain and display the results of the research experiments that validated its performance based on clustering execution, computational time, and dimensional space.

Based on four evaluation metrics, Table 2 displays how well the suggested feature selection method (PDO) performed on eight industry-standard text datasets. The evaluation measures showed that the proposed feature selection approach, which used the PDO algorithm, enhanced the text clustering performance in nearly all of the provided datasets. The suggested PDO significantly improves the text clustering method while simultaneously decreasing the feature count. In a multi-dimensional space, the suggested PDO outperforms the other comparing methods when dealing with a large collection of text documents.

Regarding feature selection performance, the suggested PDO algorithm trumped the other two comparison algorithms Gray wolf optimizer (GWO) [23] and whale optimization algorithm (WOA) [31]. The suggested PDO achieved superior results in text clustering when compared to the different comparison approach—a k-mean technique that did not use feature selection. Finally, compared to the Slime mould algorithm (SMA) [32], the proposed algorithm PDO consistently outperforms it.

Table 2 The results of comparison methods.

Dataset	**Method**	**Compared algorithms**				
		K-means	**GWO**	**WOA**	**SMA**	**PDO**
DS_1	Accuracy	.5599	.5399	.5989	.5879	.6330
	Precision	.5235	.5308	.5724	.5788	.5746
	Recall	.5111	.5800	.5715	.5552	.5736
	F-measure	.5278	.5450	.5713	.5724	.5735
DS_2	Accuracy	.3554	.3629	.4140	.4740	.4245
	Precision	.2886	.3200	.3308	.3585	.3733
	Recall	.2752	.3193	.3408	.3629	.3543
	F-measure	.391	.3184	.3420	.3593	.3620
DS_3	Accuracy	.5140	.5590	.4739	.5240	.5255
	Precision	.4755	.4645	.4296	.4820	.4886
	Recall	.4743	.4678	.4295	.4792	.4920
	F-measure	.4785	.4644	.4296	.4878	.4939
DS_4	Accuracy	.2741	.2726	.2796	.2896	.2923
	Precision	.2456	.2536	.2612	.2615	.2653
	Recall	.2548	.2533	.2513	.2660	.2640
	F-measure	.2383	.2525	.2560	.2641	.2695
DS_5	Accuracy	.4691	.4766	.4689	.4835	.4898
	Precision	.4586	.4657	.4612	.4790	.4836
	Recall	.4385	.4444	.4432	.4596	.4698
	F-measure	.4521	.4632	.4531	.4699	.4757
DS_6	Accuracy	.6269	.6469	.6515	.6682	.6695
	Precision	.6235	.6624	.6651	.6835	.6946
	Recall	.6402	.6290	.6233	.6361	.6539
	F-measure	.6366	.6424	.6486	.6542	.6719
DS_7	Accuracy	.3388	.3532	.3282	.3429	.4424
	Precision	.381	.3419	.3265	.3323	.3490
	Recall	.2946	.3295	.3235	.3192	.3298
	F-measure	.3900	.3384	.3220	.3266	.3338
DS_8	Accuracy	.4360	.4249	.4490	.4450	.4448
	Precision	.4249	.4312	.4323	.4261	.4581

Recall	.4139	.4480	.4163	.4188	.4512
F-measure	.4290	.4146	.4235	.4221	.4553

The F-measure values are used in the statistical analysis, specifically the Nemenyi test. Figure 2 displays the average rankings of the algorithms used to select text features. The proposed PDO comes out on top of the eight datasets, followed by SMA, WOA, GWO, and K-mean without feature selection.

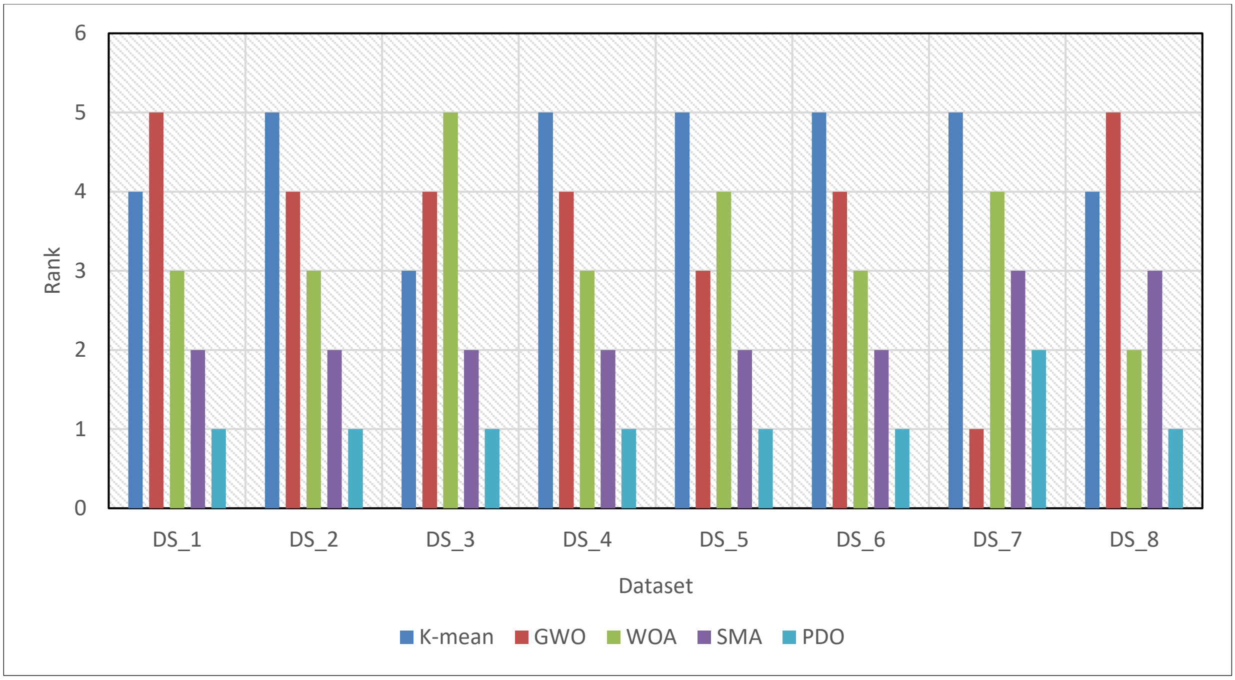

Figure 2 Average rankings of the algorithms used to select text features.

In this study, we evaluate the suggested PDO algorithm to the most effective comparative algorithms for feature selection and show the practical findings that support our claims about how well it works. In terms of F-measure performance, PDO outperformed the competing algorithms. When compared to comparable algorithms that meet all assessment criteria, it outperforms them on all eight benchmark datasets. The suggested PDO algorithm performs better when the exploration and exploitation search abilities are balanced.

5. Conclusion

In this study, we introduce PDO, a novel feature selection approach. One aspect of text document clustering is the challenge of appropriately clustering documents according to their shared attributes. The informative features will be determined before clustering so that accurate groups may be produced. For this reason, text feature selection approaches are best served by the PDO algorithm. The k-mean clustering method takes advantage of the findings from the text feature selection approach to produce reliable clusters for performance and comparison evaluations based on a text dataset related to text mining. Four text feature selection approaches are compared to find the optimal algorithm. When compared with other comparable methods, PDO produces the most favorable outcomes. Therefore, when implemented with PDO, the suggested feature selection method will lead to improved accuracy and F-measure, as well as higher-performing text features and text clustering techniques like the k-mean algorithm. Researchers in text mining may find this paper's contributions highly beneficial. Adding a new fitness function that combines the features' weighting and the number of picked features can further improve the feature selection technique. Text mining datasets that are used are well-known datasets. Nonetheless, alternative, more stringent text or date datasets can be employed for the assessment. There is another meta-heuristic approach that has been created and is working well.

Conflicts Of Interest

The author declares no conflicts of interest.

Funding

No funding.

References

[1] L. Abualigah *et al.*, "Improved prairie dog optimization algorithm by dwarf mongoose optimization algorithm for optimization problems," *Multimed. Tools Appl.*, pp. 1–41, 2023.

[2] F. Zabihi and B. Nasiri, "A novel history-driven artificial bee colony algorithm for data clustering," *Appl. Soft Comput.*, vol. 71, pp. 226–241, 2018.

[3] M. Hijjawi *et al.*, "Accelerated Arithmetic Optimization Algorithm by Cuckoo Search for Solving Engineering Design Problems," *Processes*, vol. 11, no. 5, p. 1380, 2023.

[4] O. Maâtouk, W. Ayadi, H. Bouziri, and B. Duval, "Evolutionary local search algorithm for the biclustering of gene expression data based on biological knowledge," *Appl. Soft Comput.*, vol. 104, p. 107177, 2021.

[5] M. S. Daoud *et al.*, "Recent advances of chimp optimization algorithm: Variants and applications," *J. Bionic Eng.*, vol. 20, no. 6, pp. 2840–2862, 2023.

[6] S. Xu and J. Zhang, "A parallel hybrid web document clustering algorithm and its performance study," *J. Supercomput.*, vol. 30, pp. 117–131, 2004.

[7] H. N. Alshaer, M. A. Otair, L. Abualigah, M. Alshinwan, and A. M. Khasawneh, "Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application," *Multimed. Tools Appl.*, vol. 80, no. 7, pp. 10373–10390, 2021.

[8] A.-D. Li, B. Xue, and M. Zhang, "Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies," *Appl. Soft Comput.*, vol. 106, p. 107302, 2021.

[9] M. Safaldin, M. Otair, and L. Abualigah, "Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 1559–1576, 2021.

[10] M. Ragab, "Hybrid firefly particle swarm optimisation algorithm for feature selection problems," *Expert Syst.*, 2023.

[11] L. Abualigah *et al.*, "Advances in Meta-Heuristic Optimization Algorithms in Big Data Text Clustering. Electronics 2021, 10, 101." s Note: MDPI stays neu-tral with regard to jurisdictional clai-ms in~…, 2021.

[12] N. Oikonomakou and M. Vazirgiannis, "A review of web document clustering approaches," *Data Min. Knowl. Discov. Handb.*, pp. 931–948, 2010.

[13] H.-G. Li, G.-Q. Wu, X.-G. Hu, J. Zhang, L. Li, and X. Wu, "K-means clustering with bagging and mapreduce," in *2011 44th Hawaii International Conference on System Sciences*, 2011, pp. 1–8.

[14] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Appl. Intell.*, pp. 1–39, 2022.

[15] L. Abualigah, K. H. Almotairi, M. Abd Elaziz, M. Shehab, and M. Altalhi, "Enhanced Flow Direction Arithmetic Optimization Algorithm for mathematical optimization problems with applications of data clustering," *Eng. Anal. Bound. Elem.*, vol. 138, pp. 13–29, 2022.

[16] I. B. Aydilek, "A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems," *Appl. Soft Comput.*, vol. 66, pp. 232–249, 2018.

[17] M. Q. Bashabsheh, L. Abualigah, and M. Alshinwan, "Big data analysis using hybrid meta-heuristic optimization algorithm and MapReduce framework," in *Integrating meta-heuristics and machine learning for real-world optimization problems*, Springer, 2022, pp. 181–223.

[18] L. Abualigah *et al.*, "Meta-heuristic optimization algorithms for solving real-world mechanical engineering design problems: a comprehensive survey, applications, comparative analysis, and results," *Neural Comput. Appl.*, pp. 1–30, 2022.

[19] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," *Expert Syst. Appl.*, p. 116158, 2021.

[20] O. N. Oyelade, A. E.-S. Ezugwu, T. I. A. Mohamed, and L. Abualigah, "Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm," *IEEE Access*, vol. 10, pp. 16150–16177, 2022.

[21] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Comput. Methods Appl. Mech. Eng.*, vol. 376, p. 113609, 2021.

[22] L. Abualigah, M. Shehab, M. Alshinwan, S. Mirjalili, and M. Abd Elaziz, "Ant lion optimizer: a comprehensive survey of its variants and applications," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1397–1416, 2021.

[23] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[24] L. M. Abualigah and A. T. Khader, "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering," *J. Supercomput.*, vol. 73, pp. 4773–4795, 2017.

[25] M. Mafarja *et al.*, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowledge-Based Syst.*, vol. 145, pp. 25–45, 2018.

[26] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Futur. Gener. Comput. Syst.*, vol. 97, pp. 849–872, 2019.

[27] H. Chantar, M. Mafarja, H. Alsawalqah, A. A. Heidari, I. Aljarah, and H. Faris, “Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification,” *Neural Comput. Appl.*, vol. 32, pp. 12201–12220, 2020.

[28] A. E. Ezugwu, J. O. Agushaka, L. Abualigah, S. Mirjalili, and A. H. Gandomi, “Prairie dog optimization algorithm,” *Neural Comput. Appl.*, vol. 34, no. 22, pp. 20017–20065, 2022.

[29] R. Purushothaman, S. P. Rajagopalan, and G. Dhandapani, “Hybridizing Gray Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for text feature selection and clustering,” *Appl. Soft Comput.*, vol. 96, p. 106651, 2020.

[30] L. Zhang *et al.*, “Varying the counter ion changes the kinetics, but not the final structure of colloidal gels,” *J. Colloid Interface Sci.*, vol. 463, pp. 137–144, 2016.

[31] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016.

[32] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, “Slime mould algorithm: A new method for stochastic optimization,” *Futur. Gener. Comput. Syst.*, vol. 111, pp. 300–323, 2020.

Biographies

Dr. Mohammad Alshinwan. Mohammed Al Shinwan received a Ph.D. degree from the School of Computer Engineering at Inje University, Gimhae, Republic of Korea, in 2017. He was an Assistant Professor at the Department of Computer and Information Sciences, Amman Arab University, Jordan. Currently, he is an Associate professor at an Applied science private university, in Jordan. His research interests in computer networks, Mobile networks, information security, AI, and optimization methods.. m_shinwan@asu.edu.jo

Abdul Ghafoor Memon. Associate professor at the Institute of Mathematics and Computer Science in Sindh University, Jamshoro, Pakistan. ghafoor@usindh.edu.pk

Dr Mohamed Chahine Ghanem (Chahine) is an Associate Professor and Acting Director of the Cyber Security Research Centre. He serves as the chair of the Subject Standards Board and course leader for the Digital Forensics and Cyber Security BSc (Hons) which achieved outstanding NSS and graduate outcomes since he took charge. Chahine holds an Engineering degree in Computing and Security, an MSc in Digital Foresnsics & IT Security and a PhD in Cyber Security Engineering from the City, University of London. His research focuses on applying AI to solve real cybersecurity problems and enhance automation and he published numerous research papers in top-quality journals.

Dr. Mohammed Almaayah is an Associate Professor in the Department of Computer Science at University of Jordan. Almaayah is among the top 2% scientists in the world from 2020 up to now. He is working as Editor in Chief for the International Journal of Cybersecurity and Risk Assessment. He has published over 115 research papers in highly reputed journals such as the Engineering and Science Technology, an International Journal, Education and Information Technologies, IEEE Access and others. Most of his publications were indexed under the ISI Web of Science and Scopus. His current research interests include Cybersecurity, Cybersecurity-Risk Assessment and Blockchain.