Journal of Communication Sciences and Information Technology (JCSIT)
An International Journal

Text Classification with the Whale Swarm Algorithm: A New Perspective

Amr. F. Shawish[1], Hanaa Fathi[2], Emad S. Othman[3]
[1] Lecture computer science in October Higher Institute of Engineering and Technology in 6 October city, Egypt
[2] Mathematics & Computer Science Department, Faculty of Science Menofia University
[3] Country Assoc Professor of computer and Information Systems, EL-Shorouk Academy, Cairo – Egypt.

Corresponding author: Amr. F. Shawish (Email: amr_shawish@yahoo.com).

ABSTRACT The increasing popularity of nature-inspired meta-heuristic algorithms in real-world optimization problems has been attributed to their advantages over traditional numerical optimization techniques. This study introduces the Whale Swarm Algorithm (WSA), a nature-inspired meta-heuristic, which draws inspiration from whales' communication through ultrasound for hunting. The focus of this paper is on using WSA for feature optimization. Text mining finds applications in various domains, such as business intelligence, social media analysis, sentiment analysis, biomedical analysis, software process analysis, and security analysis. This paper explores the use of WSA as an optimization algorithm, particularly in automating the understanding of Arabic text and constructing ontologies. Additionally, the paper delves into the challenges faced in enhancing the WSA models. Furthermore, this research presents a comprehensive overview of the diverse applications of WSA in different fields, with a specific emphasis on its role in ontology learning from Arabic text, aiming to improve swarm optimization techniques in practical use cases.

Keywords: optimization, swarm intelligence, whale swarm algorithm, whale optimization algorithm, ontology, Arabic text mining, concepts mapping, Association Rule Mining.

I. INTRODUCTION

Optimization is utilized in different applications. In the manufacturing of a new device, in a new artificial intelligence method, in big data application or in deep learning network, optimization is the most vital phase of any application. In order to develop a device with optimum sizes utilizing minimal power, to train a network, in order to limit the desired between the desired output and actual output values, optimization is desired. Text Mining has become an important research area. Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. Text mining can work with unstructured or semi-structured data sets such as emails, full-text documents and HTML files etc. Text Mining is widely used in field of Natural Language Processing and Multilingual Aspects. The Data Mining Optimization Ontology (DMOP) has been developed to support informed decision-making at various choice points of the data mining process. An evolutionary approach that combines information extraction technology and genetic algorithms can produce a new, integrated model for text mining. Text mining discovers unseen patterns in textual databases. The major uses of a text mining tool are for: Text Analytics, Text Processing, Classification/Categorization, Sentiment Analysis and Knowledge Discovery. Genetic Algorithms are the algorithms used to solve optimization problems. These algorithms are search based algorithm used to generate useful solutions for search problems [1].

Nature-inspired algorithms are becoming vigorous in solving numerical optimization issues, specially the NP-hard issues for example, the travelling salesman issue [3], vehicle routing [4], classification issues [5], routing issue of wireless sensor networks (WSN) [6] and multiprocessor scheduling issue [7], etc. These real world optimization issues often likely come with more global or local optima of a given mathematical model. Whereas, a point-by-point classical technique of numerical optimization is utilized for this task, the classical technique has to try repeatedly for locating various optimal solutions in every iteration [8], which takes a lot of time and work. Thus, using nature-inspired meta-heuristic algorithms to solve these issues has become an important research topic, as they are simple to execute and can converge to the global optima with high probability. In this research, we shall discuss a new nature-inspired meta-heuristic called Whale Swarm Algorithm (WSA) for function optimization; rely on the whales' behavior of communicating with every other through ultrasound for hunting. Hence, a brief overview of the nature-inspired meta-heuristic algorithms is explained. GA and WSA where merged in a new methodology, G-WSA, to extract concepts in text mining. The approach where utilized to construct Arabic ontologies from Arabic text [35]. The concepts where automatically extracted by optimizing the identification of related concepts and their relationships to parent concepts.

In this paper after the introduction in section II the natural phenomena of Whales Swarm will be retrieved, then detailed definition of the WSA is explained with the algorithm details and the mathematical background is defined in sections III and IV consequently. In section V applications of WSA is listed with related state of the art publications. Finally, a discussion and conclusion in sections VI and VII.

II. BACKGROUND

In this section different optical modulation schemes such as DCO-Q_NOMA, ACO-Q_NOMA, FLIP-Q_NOMA, ADO-Q_NOMA, and ASCO-Q_NOMA are illustrated and analyzed.

A. Whale Hunting Behavior

Social animals which live in groups inside the sea are called Whales. They make various sounds to demonstrate their movement, sustaining and mating designs. Whales decide nourishment azimuth and stay in contact with one another from enormous separations by ultrasound.

For example, pregnant females will assemble with other female whales and calves in order to improve protection abilities. Also, sperm whales are frequently seen in gatherings of somewhere in the range of 15 to 20 population, as shown in figure. 1. The whale sounds are delightful tunes in the sea and their sound range is exceptionally wide. As of not long ago, researchers have found 34 types of whale sounds, for example, whistling, squeaking, moaning, yearning, thundering, chattering, clicking, humming, churring, talking, trumpeting, clopping, etc. These sounds made by whales can frequently be connected to significant capacities, for example, their relocation, encouraging and mating designs. What's more, whales decide food azimuth and keep in contact with one another from an extraordinary separation by the ultrasound which are past the extent of human hearing [2]. system block diagram is represented as in Fig. 1.

FIGURE 1. The swarm of sperm whales.

B. Swarm Behavior

Swarming, or swarm behavior is an aggregate swarm behavior shown by creatures of comparative size which total together, maybe processing about a similar spot or maybe moving as a group or relocating toward some path. As a term, swarming is connected especially to mealy bug, likewise can be connected to some other creature that shows swarm conduct. The term rushing is typically used to allude explicitly to swarm conduct in fowls, grouping to allude to swarm conduct in quadrupeds, shoaling or tutoring to allude to swarm conduct in fish. Phytoplankton likewise assembles in tremendous swarms called blossoms, despite the fact that these living beings are green growth and are not self-impelled the manner. The term swarm is connected likewise to lifeless substances which display parallel practices, as in a robot swarm, a seismic tremor swarm, or a swarm of stars. From an increasingly theoretical perspective, swarm conduct is the aggregate movement of an enormous number of self-pushed elements. From the viewpoint of the numerical modeler, it is a developing conduct emerging from straightforward principles that are traced by people and does not include any focal coordination. Swarm conduct was first reproduced on a PC in 1986 with the reproduction program boids. This program recreates basic operators that are permitted to move as per a lot of fundamental standards. The model was initially intended to impersonate the rushing conduct of winged animals; however it tends to be connected additionally to trained fish and other swarming elements [2].

The boids PC program, made by Craig Reynolds in 1986. Many consequent and current models use minor departure from these guidelines, frequently executing them by methods for concentric "zones" around every animal. In the "zone of aversion", near the creature, the central creature will look to remove itself from its neighbors to maintain a strategic distance from impact. Second, in the "zone of alignment", the central creature will try to adjust its heading of movement to its neighbors. Third, "zone of fascination", which reaches out as far away from the central creature as it can detect, and the central creature will try to move towards a neighbor.

C. Swarm Intelligence

Swarm Intelligence (SI) is a sort of man-made consciousness that intends in order to mimic the conduct of swarms or social mealy bug. Swarm alludes to any inexactly organized accumulation of cooperating specialists. Actually swarms are viewed as decentralized self-sorted out frameworks. Swarm knowledge has a multidisciplinary character its investigation gives bits of knowledge that can enable people to oversee complex frameworks. There is no reasonable definition for swarm insight. Developing conduct, self-sorted out conduct and aggregate knowledge are the related terms. Shockingly swarm insight framework can act in a planned manner with no facilitator or outside controller.

$$x_{i\,sum}(n) = \overline{x_{odd\,\iota}(n)} + x_{even\,PC}(n) \quad (4)$$

$$x_{j\,sum}(n) = \overline{x_{odd\,J}(n)} + x_{even\,NC}(n) \quad (5)$$

III. WHALE SWARM ALGORITHM

A. Overview of Whale Swarm Algorithm

Whale swarm algorithm is developed for comprehending work enhancement issue, have romanticized a few chasing principles of whale. At the point when a whale has discovered food source, it will make sounds to tell different whale's close-by of the quality and amount of food. So every whale will get loads of warnings from the neighbors and after that transition to the legitimate spot to discover food dependent on these warnings. The behavior of whales connecting with one another through sound for hunting fill with us to build up another meta-heuristic algorithm for function optimization issues as shown figure 2.

FIGURE 4. **transmitter network diagram**

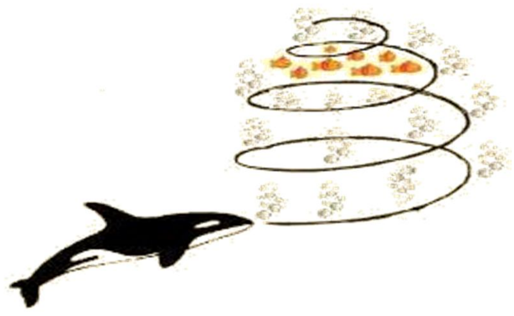

FIGURE 2. **The Method of WOA for accessing best solution.**

B. Whale Swarm Algorithm Rules

There are four rules are very important in building successful whale optimization algorithm as follows:

1. All the whales should connect with every other through ultrasound in the search region;

2. Every whale has a certain degree of computing capability in order to calculate the distance to other whales

3. The quality and quantity of food found through every whale is assigned to its fitness.

4. The motion of a whale is guided through the nearest one amongst the whales that are better (judged by fitness) than itself.

The flow chart of whale Swarm Algorithm is explained in figure 3 as follows:

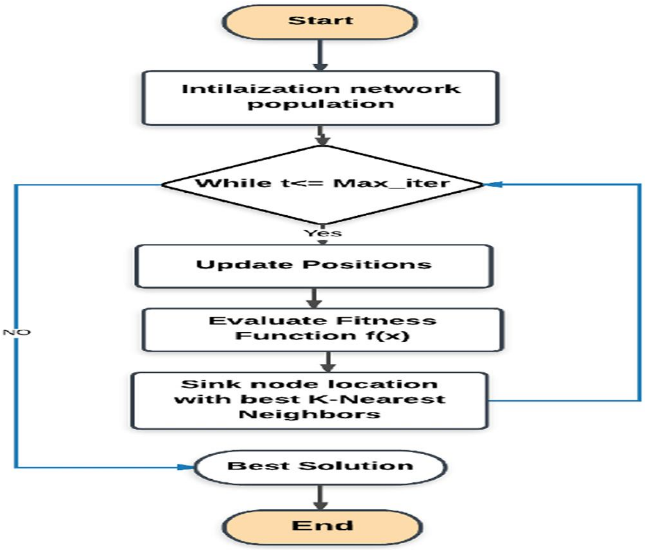

FIGURE 2. **The flow chart of WSA technique.**

IV. THE MATHEMATICAL MODEL OF WSA

The mathematical model for whale optimization algorithm is presented in details. The following functions describe the behavior of WSA to achieve encircling prey, feeding of spiral bubble-net maneuver, and search for prey.

The WOA algorithm supposes that the current best candidate solution is the target prey or is close to the optimum. After the best search agent is defined, the different search agents will hence attempt to update their positions in the direction of the best search agent. This conduct is represented with the aid of the following equations [9]

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (2)$$

Where t suggests the current iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, X* is the position vector of the best solution acquired so far, $\vec{X}$ is the position vector, | | is the absolute value, and • is an element-by-element multiplication. It is well worth citing here that X∗ should be updated in every new release if there is a better solution.

The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (4)$$

Where $\vec{a}$ is linearly reduced from 2 to 0 over the direction of iterations (in each exploration and exploitation stages) and $\vec{r}$ is a random vector in [0, 1].

For a 2D issue, the position (X, Y) of a search agent can be up to date according to the position of the current best record (X *, Y*). Various places round the best agent can be done with respect to the current position through using adjusting the value of $\vec{A}$ and $\vec{C}$ vector. It should be mentioned that with the aid of defining the random vector ($\vec{r}$) it is viable to reach any function in the search space defined among the key-points. Therefore, Eq. (2) permits any search agent to update its position in the neighborhood of the current best solution and simulates encircling the prey.

The same concept can be prolonged in order to a search space with n dimensions, and the search agents will go in hyper-cubes round the best solution got yet. As above-mentioned in the previous part, the humpback whales as well attack the prey with the bubble-net technique.

Then, Calculates the distance between the whale placed at (X, Y) and prey placed at (X*, Y*). A spiral equation is then created between the position of whale and prey to mimic the helix-shaped motion of humpback whales as follows:

$$\vec{X}(t+1) = \vec{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (5)$$

Where $\vec{D'} = |\vec{X}^*(t) - \vec{X}(t)|$ shows the distance of the i-th whale to the prey (best solution acquired yet), b is a regular for defining the shape of the logarithmic spiral, l is a random quantity in [−1, 1], and * is an element-by-element multiplication.

Humpback whales swim round the prey inside a shrinking circle and among a spiral-shaped path with each other. To mannequin this simultaneous behavior, we suggest that there is a probability of 50% to select between either the shrinking encircling mechanism or the spiral model to update the position of whales throughout optimization. The mathematical model is as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } P < 0.5 \\ \vec{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } P \geq 0.5 \end{cases} \quad (6)$$

Where p is a random number in [0, 1].

Humpback whales search randomly in accordance to the position of every other. Thus, we utilize $\vec{A}$ with the random values greater than 1 or less than −1 to force search agent to get about far away from a reference whale. In contrast to the exploitation stage, we update the position of a search agent in the exploration phase in accordance to a randomly chosen search agent instead of the best search agent found yet. This mechanism and $|\vec{A}| > 1$ confirm exploration and permit the WOA algorithm to execute a global search. The mathematical model is as follows:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}(t)| \quad (7)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (8)$$

where $\vec{X}_{rand}$ is a random position vector (a random whale) chosen from the current population.

V. USES OF WSA

There are many uses for whale swarm optimization algorithm in various areas of life. Stochastic nature-inspired meta-heuristic algorithms have validated their power on the closing two decades in dealing with global optimization issues bobbing up often in engineering. Whale optimization algorithm is a recent swarm based meta-heuristic. It copies the pattern of spiral bubble net hunting pattern of humpback. Mohit Verma and Amit Kumar (2018) [10] offered a brief survey on the whale optimization algorithm with focus on its several application over single objective and multi-objective optimization issues. First of all, the goal of single objective optimization is to attempt to uncover the exceptional solution that corresponds to the minimal or the most well worth of one objective that lumps all definitely various objectives into one. WSA has been utilized for the various single objective optimization issue. Table I lists some of the application of WSA in solving single-objective optimization issues.

A. Engineering Applications

Table 1 application of WSA in single objective optimization

S. No.	Application name	Reference
1.	Bearing fault diagnosis using a WOA	[11]
2.	Welded beam design	[12]
3.	The Workflow Planning of Construction Sites using WOA	[13]
4.	Optimal siting of capacitors in radial distribution network using WOA	[14]
5.	A hybrid WOA and pattern search technique for optimal power flow problem	[15]
6.	Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using WOA	[16]
7.	An emission constraint environment dispatch problem solution with microgrid using WOA	[17]

On the other hand, there exists no most suitable solution in case of multi-objective optimization with contradictory objectives [18]. The co-occurrence of definitely distinctive targets leads to a collection of trade-off solutions popularly known as non-dominated or Pareto-optimal solutions. Multi-Objective Whale Optimization Algorithms (MOWOA) has been utilized for the different issues with multi objective optimization. Table II lists some of the application of WSA in solving multi objective optimization issues.

Table 2 application o WOA in multi-objective optimization

S No	Application	Reference
1.	Economic and Emission Dispatch using WOA	[19]
2.	Multi-Objective Optimal Vehicle Fuel Consumption based on WOA	[20]
3.	Multi-objective optimal mobile robot path planning base on WOA	[21]
4.	An Ameliorative WOA for Multi-Objective Optimal Allocation of Water Resources	[22]
5.	A MOWOA for Solving Engineering Design Problems	[23]
6.	WOA for combined heat and power economic dispatch	[24]

B. Feature Selection Applications

Choosing of relevant benefits of a dataset is vital in high dimensional datasets to keep away from the curse of dimensionality. Feature Selection is carried out to decrease overfitting, to enhance accuracy and to decrease the training time of the algorithms. Bing Zeng, Liang Gao and Xinyu Li used WSA for feature subset selection [2]. P. Anuradha and Dr. Vasantha Kalyani David (2019) [25] focused on choosing a features subset utilizing Whale Swarm Algorithm (WSA) where Logistic Regression (LR), Random Forest (RF) and k-Nearest Neighbor (KNN) are utilized as fitness functions. These WSA-LR, WSA-RF and WSA-KNN combinations generate different feature subsets for different number of iterations. Then training and trying out is accomplished on the dataset with the subset of selected features using LR, RF, Support Vector Classifier (SVC) and Gaussian Naive Bayes (GNB) and prediction accuracies generated are analyzed.

In High dimensional datasets, Feature Selection pursuits at decreasing the redundant and irrelevant features. The refined dataset with only the relevant features would enhance the learning accuracy and decrease the learning time [25]. The features which are utilized to train the machine learning model highly influence the efficiency of the model. Irrelevant features can bring down the efficiency of the model. Feature Selection can be greatly labeled into filter technique, wrapper method and Embedded method. In Filter technique, different statistical tests are utilized to choose the features that rely on their correlation with the outcome or dependent variable. In wrapper technique, a subset of features is utilized to train a model. Based on the efficiency of the model, features will be added or deleted to/from the subset. In embedded technique, both the benefits of filter and wrapper techniques are combined. The embedded technique algorithm operates using subset selection, train a model and additionally execute a penalization function to limit overfitting.

The pseudo code of finding a whale's better and nearest whale [2]:

Input: The whale swarm n, a whale u.
Output: The better and nearest whale u.

```
1: begin
2: Define an integer variable v initialized with 0;
3: Define a float variable temp initialized with infinity;
4: for i=1 to n do
5: if i≠u then
6:       if f(whale i )
7:            if dist(whale i , whale u )
```

```
                v=i;
                temp=dist(whale i , whale u );
            end if
        end if
    end if
end for
return whale v;
end
```

The authors experimented three various classifiers for the fitness function namely, Logistic Regression, Random Forest and K-Nearest Neighbors and the subsets are obtained from various numbers of iterations. The dataset rely on the chosen subsets are then utilized for classification. The classification accuracy of four various classifiers namely, Random forest, Logistic Regression, Support Vector Classification and Gaussian Naive Bayes are compared. Among these WSA with Logistic Regression as the fitness function (WSA-LR) gives a subset of eight features on an average and the accuracy of Random Forest Classifier is found to be 85.7% which is better than the other classifiers. In future, other classifiers can be tried as fitness function in the WSA, and the prediction accuracy can be compared among other classifiers [26].

C. Clustering Applications

Clustering is a powerful method in data-mining, which entails identifying homogeneous corporations of objects based totally on the values of attributes. Meta-heuristic algorithms for example, particle swarm optimization, artificial bee colony, genetic algorithm and differential evolution are now becoming powerful techniques for clustering.

Clustering is aggregating unlabeled objects into corporations with similarities between these objects. Such that the objects in the identical clusters are extra similar to every different object in distinct clusters in accordance to some predefined criteria [27] and [28]. A variety of algorithms have been proposed that take into account the nature of the data, the volume of the information and different enter parameters in order to cluster the data. The similarity standards in clustering are a range of in different researches. Most of the clustering troubles have exponential complexity in terms of the quantity of clusters.

Lately, Mirjalili and Lewis [29] described a new swarm based meta-heuristic optimization algorithm that mimicking the social behavior of humpback whales in searching. The algorithm is inspired through the bubble net hunting delineation. They have tested the WSA algorithm with 29 mathematical benchmark optimization issues and compared the efficiency of WSA algorithm with other traditional current heuristic algorithms for example, PSO [29], Differential Evolutional [31], Gravitational Search Algorithm [32] and Fast Evolutionary Programming [33]. WSA was identified to be ample aggressive with different general and popular meta-heuristic techniques.

Jhila Nasiri1 and Farzin Modarres Khiyabani (2018) [34] proposed a new meta-heuristic clustering technique, the Whale Clustering Optimization Algorithm, primarily rely on the swarm foraging behavior of humpback whales. After a detailed formula and explanation of its implementation, they compared the proposed algorithm with different existing well-known algorithms in clustering, including PSO, ABC, GA, DE and k-Means. Proposed algorithm was once examined the usage of one synthetic and seven real benchmark data units from the UCI computer mastering repository. Simulations exhibit that the proposed algorithm can efficiently be utilized for data clustering.

The consequences of their algorithm were contrasted with customary k-means clustering strategy and other popular stochastic algorithms such as PSO, artificial bee colony, differential evolution, and genetic algorithm clustering. The Preliminary computational experience in terms of the intra-cluster distance function and standard deviation reviled that the whale optimization algorithm can successfully be applied to solve clustering issues. Furthermore, the results from the proposed algorithm was once effective, simple to execute and robust as compared with different strategies. There are some directions that can enhance the overall performance of the suggested algorithm in the future. The aggregate of WSA clustering algorithm with different clustering strategies and the usage of different fitness functions in clustering strategy should be considered in future researches.

D. Text Analysis

Rania M. Ghoniem et al. (2019) [35] proposed an optimized ontology learning from Arabic text. Ontology is a technique for extending web syntactic interoperability to semantic interoperability. Ontologies are exploited to signify massive information in such a way that permits machines to interpret its meaning, allowing it to be reused and shared. Their work was done in two phases. First, a text mining algorithm is proposed for extracting concepts and their semantic relations from text documents. The proposed algorithm calculated the *concept frequency weights* using the term frequency weights. Afterwards, they calculated the weights of thought similarity utilizing the facts of the ontology structure, involving (i) the concept's route distance, (ii) the concept's distribution layer, and (iii) the mutual mother or father concept's distribution layer. Then, feature mapping is carried out via assigning the concepts' similarities to the concept features. The second phase, a hybrid genetic-whale optimization algorithm was once proposed to optimize ontology learning from Arabic text. The operator of the G-WOA is a hybrid operator integrating GA's mutation, crossover, and selection processes with the WSA's procedures to achieve the stability between both exploitation and exploration, and to locate the solutions that showcase the best possible fitness. For estimating the overall performance of the ontology learning method, widespread comparisons are carried out utilizing extraordinary Arabic corpora and bio-inspired optimization algorithms. Moreover, two publicly accessible non-Arabic corpora are utilized to compare the performance of the proposed method with those of different languages. To validate the performance of the proposed G-WOA algorithm in mastering ontology from Arabic text, they compared the solution returned to those returned through the normal GA and WSA. The G- WSA starts off evolved to search for the fine answer via a set of iterations, which include embedding the genetic operators into the WSA architecture. Eventually, the algorithm returns the answer which recommends the great set of concepts/relations that can contribute to the ontology. The outputs revealed that the proposed genetic-whale optimization algorithm outperforms contrasting algorithms throughout all the Arabic corpora in precision, recall, and F-score measures. Moreover, the proposed approach outperforms the latest strategies of ontology learning from Arabic and non-Arabic texts in terms of these three measures.

This research contributes to todays Arabic ontology learning by the following: text mining algorithm is proposed specifically for extracting the ideas and their semantic relations from the Arabic documents. The extracted set of principles with the semantic relations constitutes the shape of the ontology. In this regard, the algorithm operated on the Arabic documents with the aid of calculating the concept frequency weights depending on the term frequency weights. Thereafter, they calculated the weights of concept similarity, using the information-driven from the ontology structure involving the concept's path distance, the concept's distribution layer, and the mutual parent concept's distribution layer. Eventually, it performs the mapping of aspects by means of assigning the notion similarity to the thinking features.

This study benefit from a prior knowledge (initial concept set obtained from the text mining algorithm) to create progressive solutions for the fine concept/relation set that can constitute the ontology. Proposed ontology learning strategy is applicable on different languages; it can be utilized to extract the most appropriate ontology structure from the non-Arabic texts. The proposed algorithm extracts standards and their semantic members of the family that constitute the ontology from every record of Arabic text, in three steps: term weighting, notion similarity weights, and feature mapping. Genetic algorithms (GA) were embedded into the WSA algorithm in order to improve a wide variety of whales (search agents) in the form of chromosomes.

The evaluation to the model was composed of three experiments: (i) comparisons with different bio-inspired optimization algorithms existing in the literature involving Arabic ontology learning, (ii) comparisons with previous published approaches on Arabic ontology gaining knowledge from text, and (ii) comparisons with modern day on gaining knowledge of ontology from non-Arabic settings. Eventually, the proposed ontology getting to know strategy is relevant to the non-Arabic texts too. It achieved higher performance that outperformed the contemporary processes on gaining knowledge of ontology from Arabic and non-Arabic text [36].

KE. Heraguemi, et.al(2021) the Whale Optimization algorithm strikes a balance between intensification and diversification, a characteristic that is also present in their proposed approach. They conducted a series of tests on well-known benchmarks and found that their method outperforms other swarm-inspired algorithms in terms of quality, runtime, and memory usage. In their paper, they focus on Association Rule Mining (ARM), which is a widely used technique for identifying relationships between items in databases. However, ARM is an NP-complete problem, and as the number of transactions and items in the database increases, the time and memory required for rule extraction become prohibitively high for exact algorithms. To address this challenge, the authors introduced the Whale Optimization Algorithm for Association Rule Mining (WO-ARM). Their approach enhances the original whale optimization algorithm by incorporating mechanisms like the shrinking encircling technique, the spiral-shaped path, and the search prey technique. They evaluated WO-ARM using six well-known datasets in the ARM field and compared the results with recently developed similar approaches. The findings demonstrated that WO-ARM is more effective in terms of runtime, quality, and memory consumption, primarily due to the mechanisms employed by the whale optimization algorithm. The authors intend to further enhance their approach to handle large-scale datasets by implementing parallel execution on Graphical Processing Units (GPU) in the near future [37].

VI. Discussion

The Whale Swarm Algorithm, if it were to exist, could be a variant of the Whale Optimization Algorithm (WOA), which is a real optimization algorithm inspired by the hunting behavior of humpback whales. In WOA, each whale represents a potential solution to the optimization problem, and the whales communicate and cooperate to improve their solutions over time. If we extend this concept to a swarm of whales, the algorithm might involve multiple subgroups of whales, each representing a potential solution. These subgroups would communicate and exchange information to collectively search for better solutions to the given problem.

Whale swarm Optimization Algorithm (WSA) is a recent swarm intelligence based meta-heuristic optimization algorithm, which simulates the natural behavior of bubble-net hunting technique of humpback whales and has been correctly applied to clear up complicated optimization troubles in a huge variety of disciplines. Therefore, when applied to large size issues, its efficiency and performance degrades due to the huge computational work load required. Distributed computing is one of the effective ways to improve the scalability of WSA for resolving large-scale issues. Whale swarm algorithm can be applied to solve nearly any optimization issues. Whale swarm Optimization Algorithm (WSA) is one of the newly proposed algorithms belonging to the class of swarm intelligence. The humpback whale is simulated so as to find optimum solutions to various optimization issues. Swarm intelligence algorithms tend to be robust against noise or outliers in the data. In text mining, where textual data can be noisy due to spelling errors, abbreviations, or inconsistent formatting, a robust algorithm can handle such challenges effectively. The algorithm could be potentially useful in topic modeling and clustering tasks, where it can identify coherent topics or groups of related documents from a vast collection. Swarm intelligence algorithms are generally adaptable and can handle dynamic environments or data changes. In text mining, this could be beneficial when analyzing streams of incoming textual data or when dealing with text from diverse sources. Swarm intelligence algorithms can often be parallelized, allowing for faster processing of data. In text mining, where dealing with large corpora of documents can be time-consuming, the parallel nature of the algorithm could expedite the analysis process. However, it's important to reiterate that as of my last update, there was no established Whale Swarm Algorithm for text mining. Therefore, these advantages are speculative and should be considered with caution. Always refer to the latest research and literature for up-to-date information on any new developments in swarm intelligence algorithms or their application to text mining.

There are nonetheless some shortcomings that are handy to fall into nearby top of the line or slow convergence, which want to be always improved and innovated. There are many complex optimization problems such as clustering, Hadoop MapReduce, Dynamic software rejuvenation in web…etc. So, in the future, we will improve the whale swarm algorithm to solve the previous issues.

Future research might explore how the Whale Swarm Algorithm can be adapted and specialized for specific domains such as healthcare, finance, social media, etc. Future work could focus on investigating the scalability of the Whale Swarm Algorithm for handling large-scale text mining tasks efficiently. The Whale Swarm Algorithm might be used to enhance NER models, leading to more accurate entity recognition. The Whale Swarm Algorithm could be applied to text summarization tasks, where the goal is to generate concise and informative summaries of lengthy documents or articles. Researchers could investigate how the Whale Swarm Algorithm could efficiently perform these tasks and discover meaningful patterns within the text corpus. Future work would involve refining and optimizing the Whale Swarm Algorithm specifically for text mining tasks. This might entail adapting the algorithm's parameters, communication patterns, and exploration strategies to suit the unique characteristics of textual data.

VII. CONCLUSION

In this paper, new swarm intelligence based meta-heuristic called Whale Swarm Algorithm, inspired using the whales' behavior of communicating with every other through ultrasound for hunting, is explained in this paper. We showed the methodology and strategy of whale swarm algorithm in detail. Finally, we explained several uses of WSA in many problems such as Clustering, Prediction Diseases, Mechanical, Text Mining and Production Engineering. Semantic understanding of textual knowledge to find concepts of thought could be detect by the use of genetics and WSA. This will provide a new layer of knowledge understanding through an optimization mechanism for big data analytics. It is still required to find means of advanced processing of such optimization technique to enhance the performance of WSA and text mining of huge amount of documents.

REFERENCES

[1] S.M. Khalessizadeh, R.Zaefarian, World Academy of Science, Engineering and Technology, 2006, "Genetic Mining: Using Genetic Algorithm for Topic based on Concept Distribution"

[2] Bing Zeng. Liang Gao. Xinyu Li. 2017. Whale swarm algorithm for function optimization. LNCS, 10361:624-639

[3] M. Mahi, Ö.K. Baykan, H. Kodaz, A new hybrid method based on Particle Swarm Optimization, Ant Colony Optimization and 3-Opt algorithms for Traveling Salesman Problem, Applied Soft Computing, 30 (2015) 484-490.

[4] K.C. Tan, Y. Chew, L.H. Lee, A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems, European Journal of Operational Research, 172 (2006) 855-885.

[5] S.N. Qasem, S.M. Shamsuddin, S.Z.M. Hashim, M. Darus, E. AlShammari, Memetic multiobjective particle swarm optimization-based radial basis function network for classification problems, Information Sciences, 239 (2013) 165-190.

[6] B. Zeng, Y. Dong, An improved harmony search based energyefficient routing algorithm for wireless sensor networks, Applied Soft Computing, 41 (2016) 135-147.

[7] E.S. Hou, N. Ansari, H. Ren, A genetic algorithm for multiprocessor scheduling, IEEE Transactions on Parallel and Distributed Systems , 5(2) , Feb 1994.

[8] B.-Y. Qu, P. Suganthan, S. Das, A distance-based locally informed particle swarm model for multimodal optimization, Evolutionary Computation, IEEE Transactions on, 17 (2013) 387-402.

[9] Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. Cogent Mathematics & Statistics, 5(1), 1483565.

[10] Mohit Verma, Amit Kumar," A Brief Review of Applications of Whale Optimization Algorithm to Mechanical and Produkction Engineering", in: International Journal of Pure and Applied Mathematics, 119(18), pp. 1953-1960, 2018.

[11] Zhang, X., Liu, Z., Miao, Q. and Wang, L., Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time– frequency

atom dictionary, Mech. Syst. Signal Process., vol. 107, pp. 29–42.
[12] Mirjalili, S. and Lewis, A., 2016. The Whale Optimization Algorithm, Adv. Eng.Softw., vol. 95, pp. 51–67.
[13] Rohani and Mohammad. 2016. The Workflow Planning of Construction Sites using whale optimization algorithm (WOA), Turkish Online Journal of Design Art and Communication 6: 2938-2950.
[14] [14] Prakash, D.B. and C.Lakshminarayana, 2016. Optimal siting of capacitors in radial distribution network using Whale Optimization Algorithm. Alexandria Engineering Journal (2017)56,499-509.
[15] Bentouati, B., Chaib, L. and Chettih, L., 2016. A hybrid whale algorithm and pattern search technique for optimal power flow problem, 2016 8th International Conference on Modelling, Identification and Control (ICMIC), IEEE 2016.
[16] Hitartch Buch, Indrajit N. Trivedi, Narottam Jangir, R.H.Bhesdadiya, Pradeep Jangir and Arvind Kumar, Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using Whale Optimization Algorithm, Conference: PIINCON-2016, At Bikaner India, Volume: https://www.piicon2016.com/index.aspx
[17] Trivedi, Indrajit N., 2016. An emission constraint environment dispatch problem solution with microgrid using Whale Optimization Algorithm, Power Systems Conference (NPSC), 2016 National. IEEE, 2016.
[18] Jangir, P. and Jangir, N., 2017. Non-Dominated Sorting Whale Optimization Algorithm (NSWOA): A Multi-Objective Optimization Algorithm for Solving Engineering Design Problems, Global Journal of Researches in Engineering: F Electrical and Electronics Engineering Volume 17 Issue 4 Version 1.0.
[19] Faseela, C.K. and Vennila, H., 2018. Economic and Emission Dispatch using Whale Optimization Algorithm (WOA), 2018 in IJECE (Vol 8, No 3). 2018.
[20] Horng. andMong, Fong., 2016. A Multi-Objective Optimal Vehicle Fuel Consumption Based on Whale Optimization Algorithm, Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2016, Kaohsiung, Taiwan, (Volume 2. Springer International Publishing). 2017.
[21] Dao, Thi-Kien, Tien-Szu Pan, and Jeng-Shyang Pan., 2016 A multi-objective optimal mobile robot path planning based on whale optimization algorithm. 2016 IEEE 13th International Conference on Signal Processing (ICSP), Chengdu, China.
[22] Yan, Z., Sha, J., Liu, B., Tian, W. and Lu, J., 2018. An Ameliorative Whale Optimization Algorithm for Multi-Objective Optimal Allocation of Water Resources in Handan, China, water, 10,87.
[23] Marler, R. T. and Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, vol. 26, no. 6. pp. 369–395.
[24] Kalaipriyan, T., Amudhavel, J. and Sujatha, P., 2017. Whale Optimization Algorithm for combined heat and power economic dispatch, Advances and Applications in Mathematical Sciences Volume 17, Issue 1, November 2017, Pages 197-211.
[25] Jie Cai. Jiawei Luo. Shulin Wang. Sheng Yang. 2018. Feature selection in machine learning: A new perspective. Neurocomputing, 300:70–79.
[26] Anuradha, P. and Dr. Vasantha Kalyani David," International Journal of Research and Analytical Reviews (IJRAR), 6(2), 2018.
[27] Elhag, A., & Ozcan, E. (2018). Data clustering using grouping hyper-heuristics. Evolutionary Computation in Combinatorial Optimization, LNCS, 10782, 101–115.
[28] Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. Experiments System Applications, 37, 4761–4767. doi: 10.1016/j. eswa.2009.11.003
[29] Mirjalili, S., & Lewi, A. (2016). The whale optimization algorithm. Advancement Engineering Softwares, 95, 51–67. doi:10.1016/j.advengsoft.2016.01.008
[30] Kenedy, J., & Eberhart, R. (1995). Particle swarm optimization. in: Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, Australia,1995.
[31] Storn, R., & Price, K. (1997). Differential evolution- a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization,11, 341–359. doi:10.1023/A:1008202821328
[32] Rashedi, E., Nezamabadi- Pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algoritm. Information Sciences, 179, 2232–2248. doi:10.1016/j.ins.2009.03.004
[33] Yao, X., & Liu, Y. (1999). Evolutionary programing made faster. IEEE Transactions Evolution Computer, 3, 82– 102. doi:10.1109/4235.771163
[34] Jhila Nasiri and Farzin Modarres Khiyabani," A whale optimization algorithm (WOA) approach for clustering", in: APPLIED & INTERDISCIPLINARY MATHEMATICS, 5(1),2018.
[35] Mezghanni, I.B.; Gargouri, F. CrimAr: A Criminal Arabic Ontology for a Benchmark Based Evaluation. Procedia Comput. Sci. 2017, 112, 653–662. [CrossRef]
[36] Rania M. Ghoniem , Nawal Alhelwa and Khaled Shaalan," A Novel Hybrid Genetic-Whale Optimization Model for Ontology Learning from Arabic Text", In: the journal of algorithms,12(9),2019.
[37] Kamel Eddine Heraguemi, Houria Kadrii and Amira Zabi, " Whale Optimization Algorithm for Solving Association Rule Mining Issue", in: nternational Journal of Computing and Digital Systems,10(1), pp. 332-342,2021.