An International Journal of Optimization and Control: Theories & Applications
ISSN: 2146-0957 eISSN: 2146-5703
Vol.11, No.2, pp.216-226 (2021)
https://doi.org/10.11121/ijocta.01.2021.001091

RESEARCH ARTICLE

An application of the whale optimization algorithm with Levy flight strategy for clustering of medical datasets

Ayşe Nagehan Mat, Onur İnan and Murat Karakoyun *

Department of Computer Engineering, Necmettin Erbakan University, Turkey
nagehanmat@gmail.com, oinan@erbakan.edu.tr, mkarakoyun@erbakan.edu.tr

ARTICLE INFO

Article history:
Received: 4 March 2021
Accepted: 25 May 2021
Available Online: 22 June 2021

Keywords:
Clustering
Whale optimization algorithm
Levy flight
K-means
K-medoids
Fuzzy c-means

AMS Classification 2010:
68T01; 68T05; 68T20

ABSTRACT

Clustering, which is handled by many researchers, is separating data into clusters without supervision. In clustering, the data are grouped using similarities or differences between them. Many traditional and heuristic algorithms are used in clustering problems and new techniques continue to be developed today. In this study, a new and effective clustering algorithm was developed by using the Whale Optimization Algorithm (WOA) and Levy flight (LF) strategy that imitates the hunting behavior of whales. With the developed WOA-LF algorithm, clustering was performed using ten medical datasets taken from the UCI Machine Learning Repository database. The clustering performance of the WOA-LF was compared with the performance of k-means, k-medoids, fuzzy c-means and the original WOA clustering algorithms. Application results showed that WOA-LF has more successful clustering performance in general and can be used as an alternative algorithm in clustering problems.

(cc) BY

1. Introduction

Due to the increasingly widespread digitalization processes at global and local level, large-scale data are obtained in many different fields. It is very important to process the data accurately and quickly in order to make more effective use of the large-scale data and extract meaningful information. For this reason, new methods are constantly being developed for the efficient use of knowledge in many areas such as industry, banking, marketing, medicine, engineering and economics, where data mining techniques are being applied.

With the ever-developing science and technology, optimization problems are also increasing. These problems are considered to be more complex and difficult because they are large-scale and have many factors. Existing optimization algorithms can be low-performance and slow in solving complex optimization problems. For this reason, many researchers have focused on improving existing optimization algorithms. Hybridization is one of these improvement efforts, and it means using two or more algorithms as a hybrid. The purpose of hybridization is to use the advantages of each algorithm at the highest level and to minimize the disadvantages. When compared with its components a hybrid algorithm generally has a stronger and robust structure and can effectively solve complex optimization problems [1-3].

Data mining reveals hidden patterns and relationships in the data by using advanced analysis techniques such as machine learning, artificial intelligence, and statistics. Clustering, which is one of the data mining techniques, collects data of similar characteristics together and ensures the data community to be divided into clusters / groups. Looking at literature, the use of heuristic algorithms in clustering emerges as an alternative to the traditional clustering techniques [4, 5].

Selim and El-Sultan [6], used simulated annealing approach for clustering problem. The predetermined parameters of the algorithm are discussed and shown to converge to the global solution in the clustering problem. Mualik and Mukhopadhyay [7] presented a unified clustering algorithm. They combined the simulated annealing algorithm with artificial neural networks to improve solution quality. The proposed hybrid algorithm was used to cluster three true microarray datasets, and the results of the proposed approach were compared with some commonly used clustering algorithms. The results showed the superiority of the new algorithm. Mualik and Bundyopadyay [8] presented a genetic algorithm-based

*Corresponding author: mkarakoyun@erbakan.edu.tr

approach to solve the cluster problem. They evaluated the performance of the approach using synthetic and real datasets. Shelokar et al. [9] proposed a clustering algorithm based on ant colony optimization (ACO). The proposed algorithm has been tested on some artificial and real datasets. The performance of this technique was promising compared to popular algorithms such as genetic algorithm, simulated annealing and Tabu search. Merwe et al. [10] used the particle swarm optimization (PSO) algorithm to solve the clustering problem. They used PSO clustering, where the swarm particles were selected by the k-means algorithm, and a hybrid method. Both methods were compared with the k-means algorithm and the proposed algorithms were observed to have better results. Tunchan [11] introduced a new PSO approach that is effective in clustering problem, easy to adjust, applicable in cases where cluster number is known or unknown. Karaboğa et al. [12] used the artificial bee colony algorithm (ABC) to solve the clustering problem. The results on the test data showed that the proposed algorithm was superior performance compared to the PSO algorithm and some other approaches. Additionally, the authors found that the ABC algorithm may be suitable for solving multivariate clustering problem. Zhang et al. [5] proposed an artificial bee colony (ABC) clustering algorithm. In this algorithm, Deb's rule was used instead of greedy approach in the selection process. The algorithm tested in several well-known datasets was compared with other popular heuristic scanning algorithms in clustering. The results were successful in terms of the quality of the clusters. Armando and Farmani [13] proposed a method that uses k-means and ABC algorithms together. In the proposed algorithm ABC algorithm assist to increase the efficiency of the k-means algorithm in finding the global optimum solution. Karthikeyan and Christopher [14] proposed an algorithm with a combination of PSO and ABC algorithms. The performance of the proposed approach was determined by comparison with other clustering algorithms. Sandeep and Pankaj [15] proposed a new hybrid sequential clustering approach that uses PSO and fuzzy k-means algorithms sequentially in data clustering. Experimental results show that the new approach improves the quality of the generated clusters and avoids local minima.

Recently, Mirjalili and Lewis [16] introduced a new metaheuristic optimization algorithm called the whale optimization algorithm (WOA) that imitates the bubble-hunting strategy of humpback whales. The WOA algorithm was tested with 29 mathematical comparison optimization problems, and the performance of the algorithm was compared to other metaheuristic algorithms such as PSO [17], Differential Evolution [18], Gravitational Search Algorithm [19] and Fast Evolutionary Programming [20]. As a result of comparisons, WOA was accepted to be able to compete with other well-known metaheuristic methods. Nasiri and Khiyabani [21] used WOA in clustering in their work. They compared the results with other popular algorithms such as k-means, PSO, artificial bee colony, differential evolution and genetic algorithm. The intra-cluster distance function and standard deviation values showed that the whale optimization algorithm could be successfully applied in solving the clustering problem. Canayaz and Özdağ [22] obtained a feature vector called BEST in their proposed method by applying clustering with WOA. They applied clustering by calculating the Manhattan distance between this vector and the test data. The results showed that WOA can be used in clustering problem and is faster than the artificial atom algorithm when evaluated in terms of running time.

Levy flight is a random walking class. It can be defined as a generalized Brownian motion to include non-Gaussian randomly distributed step lengths for moving distance [23]. Levy flight can depict many natural and artificial conditions such as liquid dynamics, earthquake analysis, diffusion of fluorescent molecules, cooling behavior, noise [24]. Levy flight was also used by Pereyra and Hajj [25] in skin tissue by ultrasound and Al-teemy [23] for Ladar (Laser Detection and Ranging) screening. In addition to these fields, it has played an important role in many fields in computer science. Internet traffic models by Terdik and Gyres [26], delay and interruption tolerance network by Chen [24], Sutantyo et al. [27] has been used by the multi-robot search procedure and by Rhee [28] in areas of human mobility. In addition, Levy flight, which is similar to the food search behavior of many animals such as albatross, wasps and deer, has been used in conjunction with natural algorithms to improve the performance of algorithms [29, 30]. Yang and Deb [31] used the Levy flight distribution to create a new cuckoo in Cuckoo Search. In addition, Yang [32] introduced a new Levy flight-Firefly algorithm (LFA) by combining the firefly algorithm (FA) with the Levy flight search strategy to improve randomness.

In this article, the advantages of the whale optimization algorithm such as having a small number of parameters and avoiding local minima trap, have been effective in selecting it for the clustering problem. The aim of the study is to achieve better results with simple solutions and to cluster unlabeled data. It has been used in conjunction with the Levy flight search strategy to strengthen WOA's global search and perform a complete search compared to existing methods. The proposed algorithm has been tested on ten medical datasets selected from the UCI [33] database. Clustering performance of the proposed algorithm is given comparatively with k-means, k-medoids and fuzzy c-means algorithms.

The rest of this paper organized as follow: Clustering problem is presented with details in section 2 and the algorithms used in the study are mentioned with details in section 3. Experimental application is explained in section 4 and experimental results are given comparatively in section 5. Finally, conclusions and some future research direction are in section 6.

2. Clustering problem

Clustering, which is one of the data mining applications, is based on grouping the elements in the data collection according to their similarities (or dissimilarities). Clustering methods help to divide the data that is not known to which group it belongs to sub-sections according to similar characteristics [34, 35]. The main purpose of clustering is to group objects using their characteristic properties. While clustering, attention is paid to cluster the individuals in the same group similar to each other, and individuals from different groups to be in a separate group. As seen in Figure 1, it wants to be that the individuals within the same cluster have little distance and different clusters have more distance from each other [36].

Clustering is done in a population where precise information about grouping of variables in the dataset is not available. Observation results of n data taken from this population are regarded for p variables. In clustering, individuals of similar nature are combined and divided into clusters. Clustering allows gathering observation results with little loss [37].

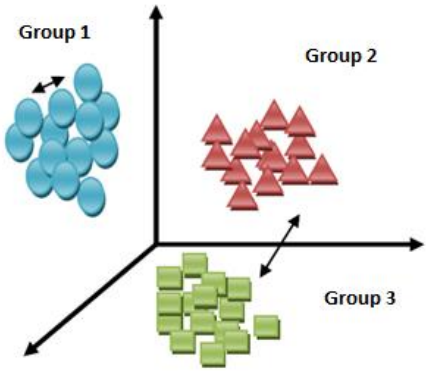


Figure 1. Distance between objects and clusters.

While there are many approaches to the clustering problem in the literature, clustering algorithms are mainly divided into two as hierarchical and non-hierarchical clustering algorithms. When applying hierarchical clustering methods, the cluster number does not need to be determined. In these methods, the cluster number is determined after the clustering process is finished. However, in non-hierarchical clustering methods, cluster number information is required to make clustering. In summary, the purpose of non-hierarchical clustering is to divide n data samples into k sets. In terms of time complexity, it is seen that hierarchical clustering methods are quadratic, while non-hierarchical clustering methods are linear [38, 39]. Clustering algorithms can be categorized as given in Figure 2.

When performing cluster analysis, a similarity or distance criterion should be selected in the first stage. Then, which clustering technique (such as hierarchical or non-hierarchical) should be determined. In the next step, the type of clustering method to be used for the selected technique is chosen and in the last stage, the number of clusters is determined in order to interpret the cluster result [41]. In this study, the distance between each data sample and the cluster center to which the data sample belongs was calculated according to the Euclidean distance [42].

3. Reflective process

Clustering has been applied to ten medical datasets using the Whale Optimization Algorithm, which has been introduced in the literature in recent years and Levy flight search strategy. Clustering results are given in comparison with results of k-means, k-medoids and fuzzy c-means clustering algorithms.

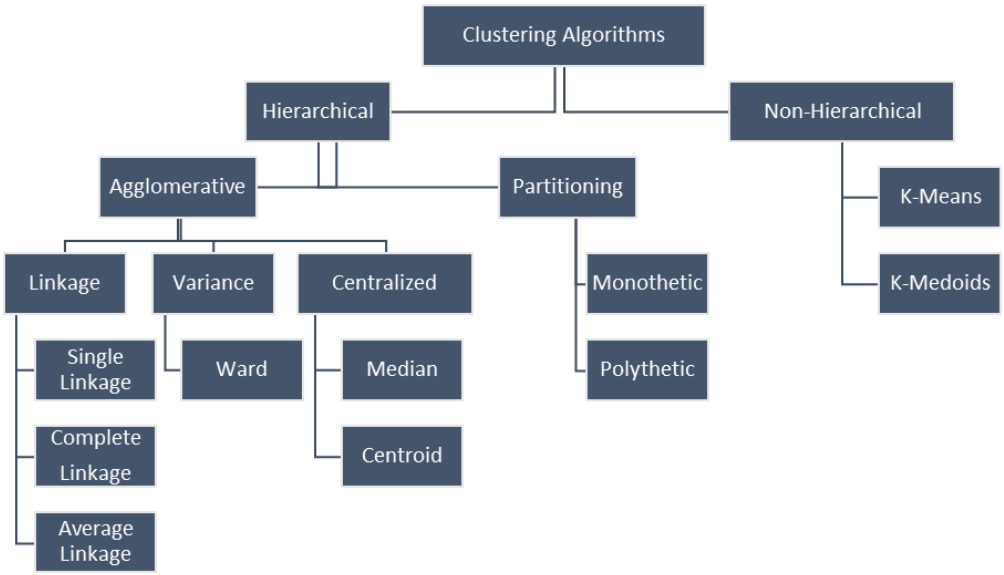


Figure 2. The general categorization of the clustering algorithms [40].

3.1. K-means clustering algorithm

Developed in 1967 by J. B. MacQueen, k-means is one of the oldest clustering algorithms [43]. One of the widely used unsupervised learning algorithms, k-means allows each object to be included in only one cluster. Therefore, it is a sharp clustering algorithm. This method is based on the idea that the cluster center represents the cluster [44].

In the evaluation of the K-means clustering method, the sum of squared errors (SSE) is often used. The lowest value of SSE represents the best result. The sum of squares of objects distance (x) from cluster centers (m_i) is calculated by the Eq. (1) [45].

$$SSE = \sum_{i=1}^{K} \sum_{x \, \epsilon \, C_i} dist^2 \ (m_i \, , x \,) \tag{1}$$

The k-means algorithm is basically based on the principle of dividing n objects into k sets. It is aimed here that the clusters are as dense as possible in terms of objects and at a maximum distance from the other clusters. Cluster similarity is measured by the average value of the objects in the cluster, which represents the center of gravity of the cluster [46].

3.2. K-medoids clustering algorithm

The k-medoids algorithm is based on finding k objects that represent various structural features of the data [47]. The representative object is the closest point to the cluster center and is called the medoid. When dividing the group of objects into k clusters, the aim is to gather similar objects together and create a structure in which the objects in different sets are unalike. Although there are many different types of the k-medoids algorithm, the first introduced k-medoids algorithm is the PAM (Partitioning Around Medoids) algorithm. PAM, similar to the k-means algorithm, determines the randomly chosen k numbers as the cluster center. With each new element joining the cluster, the elements of the cluster are tested and the point that can contribute the most to the cluster development is determined. A swap operation is performed so that this determined point is the new center of the cluster and the old cluster center is the ordinary cluster element [48].

3.3. Fuzzy c-means clustering algorithm

Fuzzy C-Means (FCM) algorithm can be called fuzzy version of the K-means clustering algorithm [49]. FCM algorithm is the best known and widely used method among fuzzy partitioning clustering techniques. It was proposed by Dunn in 1973 and was developed by Bezdek in 1981 [50]. Fuzzy c-means algorithm works with purpose function logic. Objects have a membership value in the range [0, 1] for each cluster, and according to fuzzy logic they belong to that cluster depending on these membership values. The sum of the membership values of an object calculated for all clusters must be 1. The cluster with the highest membership value refers to the cluster center where the object is closest. Clustering is completed when the fitness function converge to the specified value.

The algorithm uses it to minimize the objective function in Eq. (2), which is the generalization of the least squares method.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_{i-} \, c_j \right\|^2 , 1 \leq m < \infty \tag{2}$$

The algorithm is started with the randomly generated U membership matrix. In the second step, center vectors are calculated using Eq. (3) [50].

$$c_{j=} \frac{\sum_{i=1}^{N} u_{ij}^{m} \, x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{3}$$

The U matrix is reconstructed with the help of Eq. (4) according to the calculated cluster centers. The previous state of the U matrix is compared with the final state, and the operations are repeated until the difference is less than ε [51].

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_i \, - \, c_i \|}{\| x_i - c_k \|} \right)^{2/(m-1)}} \tag{4}$$

When clustering is complete, the fuzzy values in the U membership matrix show the clustering result. When these values are defuzzified, their equivalents can be found as 0 and 1.

3.4. Whale optimization algorithm

Metaheuristic optimization algorithms have been used frequently in engineering applications in recent years. These algorithms have several advantages, such as easy conversion, non-derivative structures, and avoiding the local minima. In addition, they have a wide range of uses, as they can produce solutions to problems in different areas.

The whale optimization algorithm is a metaheuristic optimization algorithm. The algorithm imitates the hunting behavior of humpback whales and it is inspired by the bubble hunting strategy they used during hunting [16].

Humpback whales, which usually feed on small fish flocks, can create air bubble clouds by exhaling under water with their unique air bubble behavior. These large air bubble clouds, which are interconnected, are very effective at gathering prey. Then the whale starts to rise towards the surface in the bubbles formed. As it rises, it also continues to create bubbles, and as it gets closer to its prey, it narrows the bubble circle and shrinks the target. This behavior is useful in finding, immobilizing and taking prey by surprise, as well as making it possible for the hunter to hide from its prey [52].

Figure 3 (a) represents the hunting method with bubble strategy [16]. Figure 3 (b) is a real picture of hunting [16]. In the whale optimization algorithm, the hunting strategy is divided into three parts: encircling prey, bubble-net attacking method, and searching for prey.

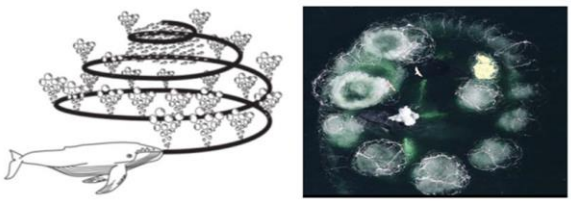


Figure 3. a) Representative hunting picture **b)** Real hunting picture.

3.4.1. Encircling prey

In the whale optimization algorithm, prey is the optimum solution to be reached. Since the optimum solution is not known in optimization problems, best result achieved or a point close to it is considered as the optimum solution. The states of other solutions are updated according to the best solution value determined. The behavior to wrap around prey is shown mathematically in Eq. (5) and (6) [16].

$$\vec{D} = \left| \vec{C} . \overrightarrow{X*}_{(t)} - \vec{X}_{(t)} \right| \tag{5}$$

$$\vec{X}_{(t+1)} = \left| \overrightarrow{X*}_{(t)} - \vec{A} . \vec{D} \right| \tag{6}$$

where t represents current iteration, $\vec{A}$ ve $\vec{C}$ convergence vectors, $\overrightarrow{X^*}$ represents the best solution vector.

$$\vec{A} = 2\,\vec{\alpha} . \vec{r} - \vec{\alpha} \tag{7}$$

$$\vec{C} = 2.\vec{r} \tag{8}$$

In Eq. (7) and Eq. (8), $\vec{r}$ represents a random vector, and $\vec{\alpha}$ represents a decreasing vector that is linear from 2 to 0 during iterations [16].

3.4.2. Bubble-net attacking method

This stage is modeled in two parts as spiral movement and narrowing the circle around the hunt. When the value of α in Eq. (7) is reduced, the circle around the prey also shrinks. Figure 4 shows the spiral motion and the position of the best solution. For the spiral movement shown, the distance between the target position (best solution candidate) and the solution candidate was calculated and Eq. (9) was created [53].

$$\vec{X}_{(t+1)} = \overrightarrow{D'} .\ e^{bl} . \cos\ (2\pi l) + \overrightarrow{X*}_{(t)} \tag{9}$$

The equation is $\overrightarrow{D'} = \overrightarrow{X*}_{(t)} - \vec{X}_{(t)}$ in Eq. (9). This expression gives the distance between the search agent and the best known position. b, refers to logarithmic spiral constant, while l refers to a random number in the range [-1, 1]. Whether the algorithm will select spiral motion or linear motion is determined by probability ½ as shown in Eq. (10). p, represents a random number in range [0, 1].

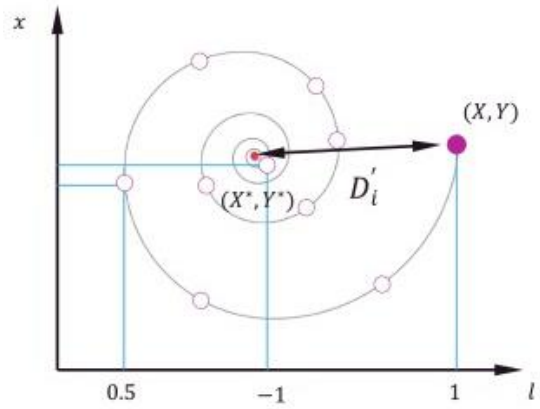


Figure 4. Spiral movement [13]

$$\vec{X}_{(t+1)} = \begin{cases} \overrightarrow{X*}_{(t)} - \vec{A} . \vec{D} , & p < 0{,}5 \\ \overrightarrow{D'} .\ e^{bl} . \cos\ (2\pi l) + \overrightarrow{X*}_{(t)} , & p \geq 0{,}5 \end{cases} \tag{10}$$

3.4.3. Search for prey

The mathematical model of the global solution is shown in Eq. (11) and (12). The new positions of candidate solutions are created around a randomly chosen candidate solution rather than the best known candidate.

$$\overrightarrow{D'} = \overrightarrow{C} . \vec{X}_{rand} - \vec{X} \tag{11}$$

$$\vec{X}_{(t+1)} = \vec{X}_{rand} - \vec{A} . \vec{D} \tag{12}$$

$\vec{X}_{rand}$, represents a randomly chosen solution vector. The value of vector $\vec{A}$ is the determinant for global or local search. A random search agent is chosen when $|\vec{A}|>1$, while the best solution is selected when $|\vec{A}|<1$ for updating the position of the search agents. The global search behavior of the algorithm is shown in Figure 5 and the pseudo code is given in Figure 6.

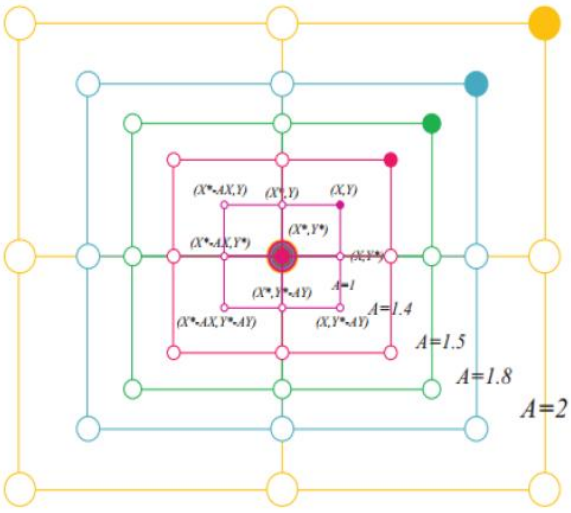


Figure 5. Global search [16].

Set initial population X_i $(i = 1,2,...,n)$
Calculate the fitness value of each search agent
X^*= the best search agent
while (t < maximum number of iterations)
for (each search agent)
Update α, A, C, l, ve p
  if (p < 0.5)
   if (|A| < 1)
    Update the position of the current search agent by Eq. (6)
   else if (|A| ≥ 1)
    Select a random search agent ($\vec{X}_{rand}$)
    Update the position of the current search agent by Eq. (12)
   end if
  else if (p ≥ 0.5)
   Update the position of the current search agent by Eq. (9)
  end if
end for
Check if any search agent goes beyond the search space and amend it
Calculate the fitness value of each search agent
Update X^* if there is a better solution
t = t + 1
end while
return X^*

Figure 6. Pseudo code of the WOA [53].

3.5. Levy flight

Levy flight was introduced in 1937 by the French mathematician Paul Levy. It is a statistical search strategy that was discovered a century or so before him and goes beyond the more traditional Brownian movement. Each of generation starts from the best known position using a series of objects and a new generation is produced at randomly distributed distances. Then, the next generation is evaluated to select the most promising one, and this process is repeated iteratively until certain stopping criterion is met.

In general, animals' foraging behavior is a kind of random movement. Since the next move depends on the current position and the possibility of moving to the next position, every random move made is of great importance. Recent studies show that Levy flight is one of the best search strategies in random motion model [54-56].

When the previous studies are examined, it is seen that Levy flight is used in applications with the modified version as well as the original version. Many researchers have used modified Levy flight techniques such as cut Levy flight, smooth cut Levy flight, and staggered cut Levy flight in optimization operations. In this study, Levy flight technique was used in its original form.

Levy flight [57] is a class of stable non-Gaussian random processes. Random motion is generated by taking advantage of the levy stable distribution. This distribution is actually a simple power law formula. Mathematically, a simple version of the Levy distribution can be defined as in Eq. (13) [31, 58]:

$$L = \begin{cases} \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\gamma}{2(s-\mu)}\right] \frac{1}{(s-\mu)^{3/2}} & if\ 0 < \mu < s < \infty \\ 0 & if\ s \le 0 \end{cases} \quad (13)$$

μ is the position or displacement parameter and γ> 0 parameter is the scale parameter that controls the distribution scale. In general, the Levy distribution should be defined as the Fourier transform.

$$F(k) = \exp\left[-\alpha|k|^{\beta}\right], 0 < \beta \le 2 \quad (14)$$

α is a parameter in the range [−1, 1] and is known as skewness or scale factor An index of stability $\beta \in (0, 2]$ is also called the Levy index. The analytic form of the integral is unknown for general β, except for a few special cases. While the parameters β and α play a major role in determining the distribution, the γ and μ parameters have a small effect. The β parameter controls the shape of the probability distribution to obtain different shapes of probability distribution, especially in the tail region. Therefore, the smaller parameter β causes the scatter to make longer jumps because it will create a longer tail [59]. The sign of the skewness parameter α indicates the skew direction of the curve. Positive values represent the right direction and negative values represent the left direction. When α = 0, the distribution is symmetrical. The last two parameters, γ width and μ change, are the peaks of the distribution [23]. Different values of the β parameter change the distribution. It makes longer jumps for smaller values, while it makes shorter jumps for larger values.

4. Experimental Application

Datasets with different number of features and number of clusters were selected to compare the performance of the clustering algorithms used in the study. Table 1 contains summary information about the datasets. Algorithms were implemented on an Intel® Core ™ i5-2400 CPU @ 3.1 GHz processor, 4 GB RAM and Windows 7 (64-bit) Professional operating system.

Table 1. Properties of datasets

Dataset	#Clusters	#Attributes	#Data
Dermatology	6	34	366
Cancer-Int	2	9	699
Cancer	2	30	569
Thyroid	3	5	215
Heart	2	13	270
Spect	2	22	267
Diabetes	2	8	768
Hepatitis	2	21	155
Breast Tissue	6	9	106
Parkinson	2	22	195

Since the WOA clustering algorithm has exploration and exploitation capabilities, it can be considered as a global optimizer. The WOA algorithm starts with a series of random solutions and updates the position of search agents according to the above equations at each iteration. The adaptive change of the search vector $\vec{A}$ allows the algorithm to switch easily between exploration and exploitation. This means that some iterations focus on exploration and the rest on exploitation. The exploration ability of the algorithm is due to the position (Eq. (12)) update mechanism used. The purpose of the WOA is to find the search agent that provides the best evaluation of a particular fitness function. In the study, the WOA fitness function value calculated during the iterations was compared with the Levy flight fitness function value. If the SSE value of the position found is better after applying the Levy flight strategy, the WOA position has been updated. Thus, it was tried to achieve a better clustering result by increasing the efficiency of global search. In all algorithms, the number of iterations was selected as 1000 and the population size as 100. K-means, k-medoids and fuzzy c-means algorithms have only the maximum number of iterations as parameter. However, the WOA algorithm includes a number of parameters. The α parameter starts with a value of 2 and decreases linearly towards 0. The parameters r_1 and r_2 are randomly generated in the range [0, 1].

5. Experimental results

In application, the SSE value given in Eq. (1) is used as a fitness function for all algorithms. The purpose of the algorithms is to find the best cluster centers that minimize the SSE value. The results of the 30 runs for the algorithms are given in Table 2 while B is best, W is worst, A is average and S is standard deviation. Considering the average values, WOA-LF had the best result on eight datasets (cancer-int, thyroid, heart, spect, diabetes, hepatitis, breast tissue, parkinson). While in dermatology dataset k-means algorithm has found the best result, three algorithms (WOA-LF, WOA, k-means) have also gained the best result in cancer dataset. K-medoids and fuzzy c-means algorithms did not achieve better results in any dataset compared to other algorithms. As a result, it is clearly seen that the clustering results obtained in the original WOA algorithm are improved in the WOA-LF hybrid method. The results show that the WOA generally has better results than other three algorithms and by using the Levy flight strategy the performance of the WOA has been improved. In this case, it is possible to clearly say that the Levy flight strategy causes positive affects for the clustering problem on the WOA algorithm.

Table 3 shows the average results of the SSE value of the 30 runs for each algorithm and the average success ranking on datasets based on these values. Since there are five algorithms, the algorithms have a rank value between 1 and 5. According to the average rankings, WOA-LF has the best rank value with 1.1. The original WOA, k-means, k-medoids, fuzzy c-means algorithms have respectively obtained ranking values as 2.0, 2.6, 4.0 and 4.3. Here it has been clearly seen the positive affect of the LF on the WOA.

6. Conclusion

In this study, clustering was applied by using Whale Optimization Algorithm (WOA), which based on the whales' hunting strategy, and Levy flight (LF) search strategy. The Levy flight strategy has been used to improve the global search of WOA algorithm and enhance clustering results. The performance of WOA-LF has been examined comparatively with three basic clustering algorithms (k-means, k-medoids, fuzzy c-means) and original WOA. Compared to these algorithms, WOA-LF has been shown to produce better results overall. WOA-LF can be used as an alternative clustering algorithm due to its success in clustering problem.

Comparing the WOA-LF clustering algorithm to hybrid clustering approaches or using other fitness functions during clustering may be the subject of future research.On the other hand, the algorithm can be improved and apply on different optimization problems.

Table 2. The results of the algorithms for clustering problem

		K-means	**K-medoids**	**F.C-means**	**WOA**	**WOA-LF**
Dermatology	B	2.03E+03	2.83E+03	5.20E+03	2.48E+03	2.41E+03
	W	2.10E+03	3.07E+03	5.20E+03	2.66E+03	2.55E+03
	A	**2.06E+03**	2.96E+03	5.20E+03	2.58E+03	2.49E+03
	S	3.56E+01	8.74E+01	4.42E-11	4.77E+01	3.95E+01
Cancer-Int	B	2.99E+03	3.46E+03	3.29E+03	2.96E+03	2.96E+03
	W	2.99E+03	4.46E+03	3.29E+03	2.97E+03	2.96E+03
	A	2.99E+03	3.79E+03	3.29E+03	2.97E+03	**2.96E+03**
	S	8.03E-01	4.33E+02	3.94E-13	7.17E-01	8.75E-02
Cancer	B	1.34E+154	1.79e+308	7.62E+156	1.34E+154	1.34E+154
	W	1.34E+154	1.79e+308	7.62E+156	1.34E+154	1.34E+154
	A	**1.34E+154**	1.79e+308	7.62E+156	**1.34E+154**	**1.34E+154**
	S	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Thyroid	B	2.00E+03	2.14E+03	2.81E+03	1.89E+03	1.89E+03
	W	2.02E+03	2.40E+03	2.81E+03	2.07E+03	1.99E+03
	A	2.01E+03	2.24E+03	2.81E+03	1.97E+03	**1.92E+03**
	S	6.89E+00	1.02E+02	7.20E-11	4.39E+01	2.10E+01
Heart	B	1.07E+04	1.16E+04	1.39E+04	1.06E+04	1.06E+04
	W	1.07E+04	1.39E+04	1.39E+04	1.07E+04	1.07E+04
	A	1.07E+04	1.26E+04	1.39E+04	1.07E+04	**1.07E+04**
	S	0.00E+00	9.79E+02	2.34E-11	1.94E+01	2.25E+01
Spect	B	5.58E+02	6.34E+02	5.58E+02	5.55E+02	5.55E+02
	W	5.58E+02	6.34E+02	5.58E+02	5.58E+02	5.55E+02
	A	5.58E+02	6.34E+02	5.58E+02	5.55E+02	**5.55E+02**
	S	0.00E+00	0.00E+00	0.00E+00	6.71E-01	1.24E-05
Diabetes	B	7.46E+04	7.32E+04	7.46E+04	7.21E+04	7.21E+04
	W	7.46E+04	7.38E+04	7.46E+04	7.42E+04	7.21E+04
	A	7.46E+04	7.34E+04	7.46E+04	7.22E+04	**7.21E+04**
	S	0.00E+00	3.02E+02	0.00E+00	3.75E+02	4.92E+00
Hepatitis	B	9.97E+03	1.04E+04	1.28E+04	9.44E+03	9.44E+03
	W	1.14E+04	1.23E+04	1.28E+04	1.02E+04	9.48E+03
	A	1.09E+04	1.09E+04	1.28E+04	9.50E+03	**9.46E+03**
	S	6.99E+02	8.12E+02	1.20E-11	1.81E+02	1.33E+01
Breast Tissue	B	1.31E+05	1.73E+05	1.46E+05	1.26E+05	1.25E+05
	W	1.90E+05	3.94E+05	1.46E+05	1.35E+05	1.36E+05
	A	1.44E+05	2.75E+05	1.46E+05	1.30E+05	**1.28E+05**
	S	2.62E+04	7.99E+04	4.37E-11	2.47E+03	2.23E+03
Parkinson	B	1.71E+04	1.71E+04	1.71E+04	1.65E+04	1.65E+04
	W	1.71E+04	1.71E+04	1.71E+04	1.65E+04	1.65E+04
	A	1.71E+04	1.71E+04	1.71E+04	1.65E+04	**1.65E+04**
	S	0.00E+00	0.00E+00	0.00E+00	5.47E-01	5.49E-01

Table 3. Average ranking values of the algorithms

	K-means	K-medoids	F.C-means	WOA	WOA-LF
Dermatology	**2.06E+03**	2.96E+03	5.20E+03	2.58E+03	2.49E+03
	1	4	5	3	2
Cancer-Int	2.99E+03	3.79E+03	3.29E+03	2.97E+03	**2.96E+03**
	1	5	4	2	1
Cancer	**1.34E+154**	1.79e+308	7.62E+156	**1.34E+154**	**1.34E+154**
	1	4	5	1	1
Thyroid	2.01E+03	2.24E+03	2.81E+03	1.97E+03	**1.92E+03**
	3	4	5	2	1
Heart	1.07E+04	1.26E+04	1.39E+04	1.07E+04	**1.07E+04**
	3	4	5	2	1
Spect	5.58E+02	6.34E+02	5.58E+02	5.55E+02	**5.55E+02**
	3	4	3	2	1
Diabetes	7.46E+04	7.34E+04	7.46E+04	7.22E+04	**7.21E+04**
	4	3	4	2	1
Hepatit	1.09E+04	1.09E+04	1.28E+04	9.50E+03	**9.46E+03**
	4	3	5	2	1
Breast Tissue	1.44E+05	2.75E+05	1.46E+05	1.30E+05	**1.28E+05**
	3	5	4	2	1
Parkinson	1.71E+04	1.71E+04	1.71E+04	1.65E+04	**1.65E+04**
	3	4	3	2	1
Avg. Rank	2.6	4	4.3	2	**1.1**

References

[1] Evirgen, F. (2016). Analyze the optimal solutions of optimization problems by means of fractional gradient based system using VIM. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)*, 6(2), 75-83.

[2] Evirgen, F. (2017). Conformable fractional gradient based dynamic system for constrained optimization problem. *Special issue of the 3rd International Conference on Computational and Experimental Science and Engineering (ICCESEN 2016)*,1066-1069.

[3] Evirgen, F. and Yavuz, M. (2018). An alternative approach for nonlinear optimization problem with Caputo-Fabrizio derivative. *ITM Web of Conferences*,01009.

[4] Cui, D. (2017). Application of whale optimization algorithm in reservoir optimal operation. *Advances in Science and Technology of Water Resources*, 37(3), 72-79.

[5] Zhang, C., Ouyang, D., and Ning, J. (2010). An artificial bee colony approach for clustering. *Expert systems with applications*, 37(7), 4761-4767.

[6] Selim, S. Z. and Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern recognition*, 24(10), 1003-1008.

[7] Maulik, U. and Mukhopadhyay, A. (2010). Simulated annealing based automatic fuzzy clustering combined with ANN classification for analyzing microarray data. *Computers & operations research*, 37(8), 1369-1380.

[8] Maulik, U. and Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern recognition*, 33(9), 1455-1465.

[9] Shelokar, P., Jayaraman, V. K., and Kulkarni, B. D. (2004). An ant colony approach for clustering. *Analytica Chimica Acta*, 509(2), 187-195.

[10] Van der Merwe, D. and Engelbrecht, A. P. (2003). Data clustering using particle swarm optimization. *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*,215-220.

[11] Cura, T. (2012). A particle swarm optimization approach to clustering. *Expert Systems with Applications*, 39(1), 1582-1588.

[12] Karaboga, D. and Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied soft computing*, 11(1), 652-657.

[13] Armano, G. and Farmani, M. R. (2014). Clustering analysis with combination of artificial bee colony algorithm and k-means technique. *International Journal of Computer Theory and Engineering*, 6, 141-145.

[14] Karthikeyan, S. and Christopher, T. (2014). A hybrid clustering approach using artificial bee colony (ABC) and particle swarm optimization. *International Journal of Computer Applications*, 100(15).

[15] Mane, S. U. and Gaikwad, P. G. (2014). Hybrid particle swarm optimization (HPSO) for data clustering. *International Journal of Computer Applications*, 97(19).

[16] Mirjalili, S. and Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, 95, 51-67.

[17] Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-international conference on neural networks*,1942-1948.

[18] Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global

optimization over continuous spaces. *Journal of global optimization*, 11(4), 341-359.
[19] Rashedi, E., Nezamabadi-Pour, H., and Saryazdi, S. (2009). GSA: a gravitational search algorithm. *Information sciences*, 179(13), 2232-2248.
[20] Yao, X., Liu, Y., and Lin, G. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary computation*, 3(2), 82-102.
[21] Nasiri, J. and Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. *Cogent Mathematics & Statistics*, 5(1), 1483565.
[22] Canayaz, M. and Özdağ, R. (2017). Data clustering based on the whale optimization. *Middle East Journal of Technic*, 2(2), 178-187.
[23] Al-Temeemy, A. A., Spencer, J., and Ralph, J. (2010). Levy flights for improved ladar scanning. *2010 IEEE International Conference on Imaging Systems and Techniques*,225-228.
[24] Chen, Y. (2010). Research and simulation on Levy flight model for DTN. *2010 3rd International Congress on Image and Signal Processing*,4421-4423.
[25] Pereyra, M. A. and Batatia, H. (2010). A Levy flight model for ultrasound in skin tissues. *2010 IEEE International Ultrasonics Symposium*,2327-2331.
[26] Terdik, G. and Gyires, T. (2008). Lévy flights and fractal modeling of internet traffic. *IEEE/ACM Transactions on Networking*, 17(1), 120-129.
[27] Sutantyo, D. K., Kernbach, S., Levi, P., and Nepomnyashchikh, V. A. (2010). Multi-robot searching algorithm using Lévy flight and artificial potential field. *2010 IEEE Safety Security and Rescue Robotics*,1-6.
[28] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., and Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking*, 19(3), 630-643.
[29] Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev, V., *et al.* (2007). Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165), 1044-1048.
[30] Viswanathan, G. M., Afanasyev, V., Buldyrev, S., Murphy, E., Prince, P., and Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581), 413-415.
[31] Yang, X.-S. and Deb, S. (2013). Multiobjective cuckoo search for design optimization. *Computers & Operations Research*, 40(6), 1616-1624.
[32] Yang, X.-S. (2010). Firefly algorithm, Levy flights and global optimization. *Research and development in intelligent systems XXVI*. Springer, 209-218.
[33] Murphy, P. and Aha, D. (1994). *UCI repository of machine learning*. University of California, Department of Information and Computer Science.
[34] Ozdamar, K. (2002). *Paket programlari ile istatistiksel veri analizi-1*. Kaan Kitabevi, Eskisehir.
[35] Tatlıdil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Cem Ofset Ltd. Şti, Ankara.
[36] Hair Jr, J., Anderson, R., and Tatham, R. (1998). *Multivariate data analysis*.NJ: PrenticeYHall Inc, Upper Saddle River.
[37] Lorr, M. (1983). *Cluster analysis for social scientists*.Jossey-Bass Incorporated Pub.
[38] Boushaki, S. I., Kamel, N., and Bendjeghaba, O. (2018). A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications*, 96, 358-372.
[39] Frigui, H. and Krishnapuram, R. (1999). A robust competitive clustering algorithm with applications in computer vision. *Ieee transactions on pattern analysis and machine intelligence*, 21(5), 450-465.
[40] Karakoyun, M. (2015). *Kurbağa sıçrama algoritmasının kümeleme problemlerine uygulanması*. Master Thesis. Selçuk University.
[41] Sharma, S. (1996). Applied multivariate techniques.
[42] Tabak, J. (2014). *Geometry: the language of space and form*. Infobase Publishing.
[43] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*,281-297.
[44] Han, J. and Kamber, M. (2010). Data mining: concepts and techniques. [Nachdr.], *Amsterdam: Elsevier/Morgan Kaufmann*, 11, 6.
[45] Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Classification: basic concepts, decision trees, and model evaluation. *Introduction to data mining*, 1, 145-205.
[46] Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3), 645-678.
[47] Dodge, Y. (2012). *Statistical data analysis based on the L1-norm and related methods*. Birkhäuser.
[48] Dinçer, Ş. E. (2006). *Veri madenciliğinde K-means algoritması ve tıp alanında uygulanması*. Master Thesis. Kocaeli University.
[49] Karakoyun, M., Saglam, A., Baykan, N. A., and Altun, A. A. (2017). Non-locally color image segmentation for remote sensing images in different color spaces by using data-clustering methods. *5th International Conference on Advanced Technology & Sciences (ICAT'17)*, 6-12.
[50] Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.
[51] Moertini, V. (2002). Introduction to Five DataClustering Algorithms Clustering Algorithm. *Integral*, 7(2).
[52] Goldbogen, J. A., Friedlaender, A. S., Calambokidis, J., Mckenna, M. F., Simon, M., and Nowacek, D. P. (2013). Integrative approaches to the study of baleen whale diving behavior, feeding performance, and foraging ecology. *BioScience*, 63(2), 90-100.
[53] Tanyıldızı, E. and Cigalı, T. (2017). Kaotik Haritalı Balina Optimizasyon Algoritmaları. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 29(1), 307-317.
[54] Pavlyukevich, I. (2007). Lévy flights, non-local search and simulated annealing. *Journal of Computational Physics*, 226(2), 1830-1844.
[55] Reynolds, A. M. and Frye, M. A. (2007). Free-flight odor tracking in Drosophila is consistent with an optimal intermittent scale-free search. *PloS one*, 2(4), e354.

[56] Shlesinger, M. F. (2006). Search research. *Nature*, 443(7109), 281-282.
[57] Chechkin, A. V., Metzler, R., Klafter, J., and Gonchar, V. Y. (2008). Introduction to the theory of Lévy flights. *Anomalous transport*, 1, 129.
[58] Yang, X.-S. (2010). *Engineering optimization: an introduction with metaheuristic applications*. John Wiley & Sons.
[59] Lee, C.-Y. and Yao, X. (2001). Evolutionary algorithms with adaptive lévy mutations. *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)*, 568-575.

Ayşe Nagehan Mat *received the undergraduate degree from the Department of Computer Engineering of Selcuk University and still working for Master's degree. She is working as an International Officer at Selcuk University.*
https://orcid.org/0000-0003-4975-6418

Onur İnan *received the Ph.D. degree from the Department of Computer Engineering of Selcuk University. He is working as a Doctor lecturer at the Computer Engineering Department of Necmettin Erbakan University.*
https://orcid.org/0000-0003-4573-7025

Murat Karakoyun *received the Ph.D. degree from the Department of Computer Engineering of Konya Technical University. He is working as a Research Assistant at Computer Engineering Department of Necmettin Erbakan University.*
https://orcid.org/0000-0002-0677-9313

This work is licensed under a Creative Commons Attribution 4.0 International License. The authors retain ownership of the copyright for their article, but they allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in IJOCTA, so long as the original authors and source are credited. To see the complete license contents, please visit http://creativecommons.org/licenses/by/4.0/.