International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 22-33

http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

Clustering with the Blackwinged Kite Algorithm

Abdul Qadeer Rasooli[1*], Onur Inan[2], Sema servi[3]
[1*]Selçuk University, Faculty of Technology, Department of Computer Engineering
E-mail: rasooli.csf@gmail.com, ORCID https://orcid.org/0000-0003-1950-3209
[2]Selçuk University, Faculty of Technology, Department of Computer Engineering
E-mail: oinan@selcuk.edu.tr
[3]Selçuk University, Faculty of Technology, Department of Computer Engineering
E-mail: semaservi@selcuk.edu.tr

Corresponding author: rasooli.csf@gmail.com

Received: 2024-07-10 Accepted: 2024-09-17 Published online: 2024-10-01

Abstract

In this article, a metaheuristic optimization algorithm called the Black Kite Algorithm (BKA) is proposed, inspired by the nomadic and predatory behaviors. BKA integrates the Cauchy mutation strategy and the Leader strategy to enhance the algorithm's global search capability and convergence speed. This novel combination provides a good balance between exploring global solutions and utilizing local information. Clustering is a widely used technique in data analysis. Its fundamental purpose is to reveal structures and relationships in a dataset by grouping data points with similar characteristics. These groups can be utilized to understand patterns in the dataset, perform data exploration, and make predictions. Clustering algorithms typically work by measuring similarities between data points using a distance metric. Determining the number of clusters is a challenging task, even when clustering is done correctly. To address these challenges, several techniques have been proposed in the literature. Most of these methods require prior knowledge of the number of clusters to be addressed, which should be provided as an algorithm parameter. Real-world clustering problems arise when the number of clusters in the data collection set is unknown in advance. The Blackwinged Kite Optimization algorithm is a powerful search algorithm that has been proposed to solve optimization problems seen as clustering problems. With this developed method, datasets are divided into clusters based on distance, and the number of clusters is also accurately determined at a satisfactory level.

Keywords: Nature-inspired optimization, Blackwinged Kite algorithm, Metaheuristic, Constrained problems.

1. INTRODUCTION

Clustering is a fundamental technique used in data analysis to explore structures in datasets. This technique forms homogeneous groups by grouping together data points with similar characteristics. Its purpose is to understand patterns and relationships in the dataset, uncover hidden information within the data, and guide data-driven decision-making processes. Clustering analysis is commonly used in machine learning and data mining projects, marketing strategies, analysis of biological data, and many other fields.

Therefore, understanding the basic principles of clustering techniques is important for successfully applying them in data analysis processes.

Meta-heuristic optimization algorithms have seen rapid growth in this context, driven by their adaptability and gradient-free approaches. These algorithms have become critical tools in solving issues related to boosting production efficiency. The adaptability of meta-heuristic optimization algorithms enables them to adjust to a wide range of production environments and problem situations. They can explore and uncover the problem domain based on the specific problem's characteristics, pinpointing the optimal solution or a close approximation. Even when confronted with challenges like product design, production planning, resource allocation, or supply chain management, meta-heuristic optimization algorithms can be flexibly tailored and optimized to respond to real-world situations.

2. LITERATURE REVIEW

In recent years, data clustering has been the subject of many research projects, not only as a critical data mining activity but also as a dynamic way to evaluate the effectiveness of optimization algorithms. Among the effective methods to solve clustering problems are classical clustering algorithms such as K-means [1]. Despite the advantages of these algorithms, it can be challenging to discover the best solutions because they have a strong dependency on the initial settings. One of the main disadvantages of this algorithm is that it can get trapped in local optima. The second disadvantage is that the initial values of the cluster centers have a significant impact [2]. Many optimization techniques inspired by nature have been proposed to overcome these problems and enhance the effectiveness of data clustering.A broad spectrum of optimization algorithms are available for data clustering, as indicated by the literature. Ant Colony Optimization (ACO), which emulates the actions of real ants [3], is a probabilistic approach that can find effective solutions to optimization problems[4].

Another algorithm that imitates the hunting techniques of gray wolves is called Gray Wolf Optimization (GWO)[5]. Problems with clustering have been resolved with it[6]. Symbiotic Organism Search (SOS) is another method that has been suggested for clustering analysis[7]. It is modeled after symbiotic interactions inside a paired organism connection. Another optimization technique largely motivated by the notion of black holes is the Black Hole Algorithm (SDA)[8]. Introduced for clustering challenges, it is a kind of metaheuristic technique grounded in physical phenomena[9]. One further optimization technique that aims to imitate the krill swarm mechanism is the Krill Swarm technique (KSA)[10]. Furthermore, clustering problem solving is suggested [11]. The last clustering algorithm that was influenced by humpback whale behavior in its feeding mechanisms is the Whale Optimization Algorithm (WOA)[12]. In order to solve a variety of optimization problems, all of the aforementioned optimization methods have proven to be highly

Vol. 9, No. 1, 2024, pp. 22-33

http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

effective. This spurs us on to keep exploring this area and suggest an alternative to BWOA for data clustering.

Data mining refers to the process of finding patterns, trends, and insights in large databases through various computational methods. It requires extracting relevant information and knowledge from the data in order to be applicable in prediction and decision-making. Data mining utilizes a range of techniques such as regression, association rule mining, clustering, classification, and anomaly detection. Among these, clustering is a fundamental method for grouping related data points based on specific attributes or characteristics. The purpose of clustering techniques is to segregate the data into groups or clusters based on how similar the data points are to one another compared to those in other groups. By grouping unlabeled objects according to their commonalities, a process known as clustering is created, making the objects within a cluster more similar to one another than to those outside of it[8].

A sort of clustering technique called partitioning clustering algorithms is intended to break a dataset into distinct groups or clusters according to predetermined similarity standards. These algorithms divide data points into clusters iteratively in order to maximize or minimize a chosen objective function, like intra-cluster distance or inter-cluster distance. K-means, which assigns data points to the nearest centroid until convergence, is a well-known partitioning technique that iteratively updates cluster centers to divide the data into K clusters. Additional examples include CLARANS and K-medoids. These algorithms can handle very large datasets and are computationally efficient, but they require a predetermined number of clusters and can converge to suboptimal solutions depending on the initial cluster centers[13].

Rather than considering object proximity, density-based clustering algorithms group items into clusters based on local density conditions[14]. These methods see clusters as high-density zones separated from low-density, noisy areas. Density-based methods are resilient to noise and can detect non-convex clusters. However, density-based approaches such as hierarchical and partitioning algorithms face difficulties in high-dimensional spaces due to the inherent scarcity of feature space, which reduces any tendency towards clustering[15].

A soft clustering method called fuzzy c-means (FCM) clustering assigns each data point to each cluster according to its membership degree rather than forcing it to belong to just one cluster. In FCM, each data point contributes to the center of each cluster based on its membership degree, which is determined by how well the data point fits the cluster prototype. In order to minimize the objective function, which is often the weighted sum of squared distances between data points and cluster centers, the algorithm updates the membership degrees and cluster centers iteratively until convergence.

FCM is useful when dealing with data points that simultaneously belong to multiple clusters or when the boundaries between clusters are uncertain. However, the number of clusters needs to be specified [16].

Based on the distances between each data point and the cluster centers, K-means clustering, represented by the letter K, assigns the data points to one of K clusters. The cluster centroids in the space are first assigned at random. After that, a cluster is assigned to each data point according to how far it is from the cluster center. New cluster centers are assigned following the assignment of each point to a cluster. This method looks for a decent clustering iteratively. In this study, we'll suppose that the number of clusters is fixed and that we have to give a particular group points [16],[17].

3. BEHAVIOR OF INSPIRATION AND BLACK-WINGED HAWKS

The black-winged hawk is a tiny bird with a white lower body and a blue-gray upper body. It has prominent traits like migratory and hunting habits. Its great flying abilities let it to hunt with exceptional success. It feeds on small mammals, reptiles, birds, and insects. We have created an algorithm model based on black-winged hawks, drawing inspiration from their migration patterns and hunting techniques[18].

Figure 1a shows a black-winged hawk soaring in the air, while Figure 1b depicts a black-winged hawk swiftly running towards its prey [18].

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 22-33
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

4. AGGRESSIVE BEHAVIOR

Black-winged hawks are silent gliders who scan their prey, then drop and strike quickly. They hunt tiny grassland mammals and insects. During flight, they modify the angles of their wings and tails based on the wind speed. Several assault behaviors are used in this method for global search and exploration. A black-winged hawk is shown in Figure 1a flying through the air while keeping its balance and spreading its wings.

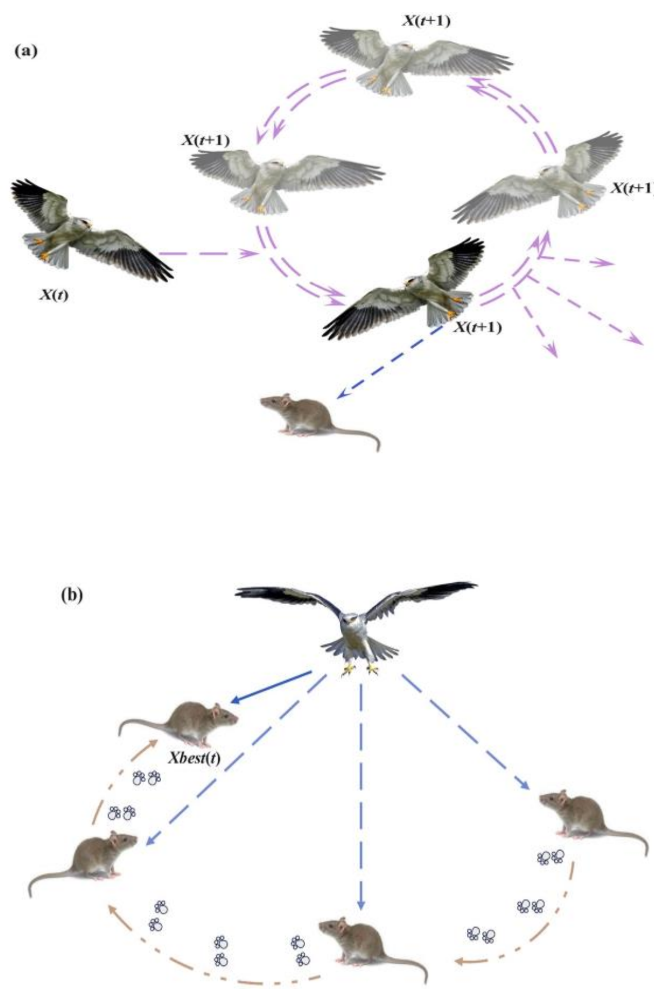

Figure 2. Two attack strategies of black-winged hawks include hovering in the air, waiting for an attack, and hovering in the air while searching for prey [18].

Figure 1b shows the scene of a black-winged hawk rapidly approaching its prey. Figure 2a illustrates the attack situation of a black-winged hawk while soaring in the air, while Figure 2b depicts the situation of a black-winged hawk while soaring in the air.

5. MIGRATION BEHAVIOR

The complicated behavior of birds during migration is controlled by various environmental conditions, including food sources and climate. In order to adjust to seasonal variations, birds migrate. During the winter, numerous birds migrate from the north to the south in search of better living circumstances and resources. Leaders typically lead flocks during migration, and their ability to navigate is essential to the flock's success.

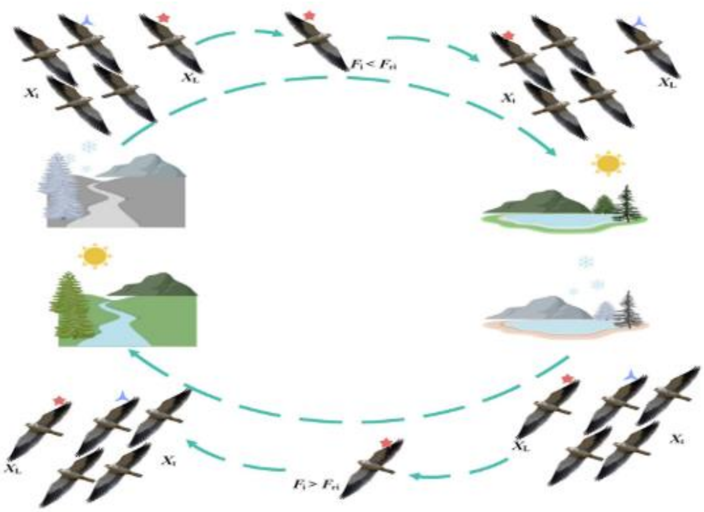

Figure 3. Black-Winged Falcons during migration[18].

6. CLUSTERING WITH THE BLACK-WINGED FALCON ALGORITHM

The Black-Winged Falcon (BWF) algorithm is a clustering algorithm inspired by a natural life form. This algorithm is developed by taking inspiration from the hunting behaviours of a predatory bird, the black-winged falcon. The black-winged falcon performs high-speed maneuvers to catch its prey and carefully scans the surrounding environment while tracking its prey. Similarly, the Black-Winged Falcon algorithm provides an effective and fast method for clustering data. This algorithm separates data points into clusters in a decentralized manner and uses natural search strategies to optimize the center and spread of each cluster. As a result, it effectively works to find data points with similar characteristics in the dataset and improves clustering performance.

The Black-Winged Falcon algorithm provides fast and accurate clustering results, especially when used with large and complex datasets. Developed by taking inspiration from nature, this algorithm offers users a valuable tool in the field of data analytics and machine learning.

Pseudocode of BWF

```
Algorithm: Black-winged kite algorithm
Input: The population size pop, maximum number of iterations T, and variable dimension dim
Output: The best quasi-optimal solution obtained by BKA for a given optimization problem.
Initialization phase
Initialization of the position of Black-winged kites and evaluation of the objective function.
Calculate the fitness value of each Black-winged kite
    while (t < T) do
/* Attacking behavior */
        if p < r
```

Figure 4. Pseudocode of BWF1

7. TUNICA SWARM ALGORITHM (TSA)

The Tunica Swarm Algorithm (TSA), an optimization method, was developed after studying the social dynamics and flocking behaviours of tunicates, marine organisms distinguished by their distinct mobility and filtering characteristics. Tunicates exhibit behaviours that optimize their positions in the water collectively to maximize nutrient absorption and minimize energy expenditure. To tackle complex optimization problems, TSA mimics these biological processes by using a population of candidate solutions (representing tunicates) that iteratively change their positions based on social interactions and individual experiences. This algorithm strikes a balance between exploration and exploitation to achieve effective convergence towards optimal solutions.

TSA has proven to be more accurate and computationally efficient than traditional optimization techniques, and it has been successfully applied to a wide range of engineering and scientific challenges. Its adaptive mechanisms, which continuously maintain a balance between local exploitation and global exploration, are what empower it. Due to its flexibility, TSA is particularly helpful in solving high-dimensional, multimodal optimization problems where other algorithms struggle to discover global optima. Further advancements in computational intelligence and biological knowledge can stem from a biologically inspired foundation that also provides insights into natural optimization processes.

8. DATA CLUSTERING APPROACH

The process of assembling N data objects into K groups according to their similarity constitutes the assembly of a data set in T-dimensional space. In order to maximize the similarity between data objects in the same group, clustering iteratively separates data objects into K groups (clusters). Furthermore, data clustering is a kind of unsupervised learning strategy where data items are clustered based on the data's structure without any training. On the other hand, data objects in supervised learning tasks, such as classifications, are categorized using labelled data and the training set. With a predetermined number of clusters, the suggested BKOA clusters data. The criterion for assessing the quality of clustering is the total of the inner cluster distances of the data objects within cluster K.

9. EXPERIMENTAL RESULTS

In this study, the performance of the RDA clustering approach was compared with well-known algorithms such as TSA clustering proposed by William H. Press and Saul A. All algorithms were programmed in Matlab 2022b and used Intel core, i7 CPU, 8 Gb and 2.6 GHz running Microsoft Windows 10. was executed on the computer. Parameter settings are the same as in the corresponding original articles. Here, 2 datasets are used to compare the performance of the proposed algorithm with the above heuristics.

The proposed work evaluates the Best, Worst, Maximum, Minimum and Average fitness values of meta-heuristic algorithms for clustering RDA.

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 22-33
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

Table 1. Comparison of the performance of the RDA clustering algorithm with the well-known clustering algorithms of TSA

Dataset	Criteria	BKA	TSA
Balance	**A**	1487.5953	1502.1398
	S	**4.4298**	5.7403
Credit	**A**	1185454.2705	1653922.1295
	S	455950.1286	**404652.759**
Dermatology	**A**	3337.3909	3410.7612
	S	55.0175	**25.2181**
Ecoli	**A**	141.5728	138.9561
	S	**7.4599**	9.3343
Glass	**A**	589.5585	687.2979
	S	66.5334	**48.4254**
Iris	**A**	191.1259	213.3086
	S	21.0529	**19.3827**
Seeds	**A**	527.3587	547.0363
	S	**29.6154**	40.2036
Thyroid	**A**	3728.0749	3929.0082
	S	**405.3393**	701.1303
Wine	**A**	19857.7439	19588.1981
	S	**729.2707**	1375.0879
Heart	**A**	12860.090	12290.6526
	S	1015.3492	**747.2681**
Spectrum	**A**	662.8352	659.8949
	S	**11.2648**	19.614
Diabetes.	**A**	95929.956	93733.439
	S	10540.7077	**7685.4059**
Hepatitis.	**A**	12611.2384	12471.3129
	S	**1263.2847**	1258.3912
Connective tissue	**A**	421054.8455	405517.1832
	S	**0.12652**	0.16799
Parkinson's disease	**A**	17774.7934	18140.9037
	S	**619.2753**	1474.7983
Somerville2	**A**	364.4995	372.8451
	S	**16.4698**	19.5627
User modeling	**A**	123.9847	122.9621
	S	**3**	3.3539

According to the results in the table above, I conducted 35 iterations on each dataset and obtained separate results for the Best, Worst, Average, and StdDev fitness values from each dataset. The local search performance of the algorithm improves significantly in the last 35 iterations, and ultimately TSA shows worse performance than BKA.

10. CONCLUSION

In summary, the utilization of the Black-winged Hawk Optimization Algorithm for Data Clustering shows promising results in solving challenging clustering problems. This method serves as an inspiration by mimicking the hunting approach of Black-winged Hawks to balance exploration and exploitation, effectively optimizing data cluster centers. By iteratively dividing the data into meaningful clusters, it contributes to enhancing data analysis and pattern recognition activities in various fields. This technique presents a competitive alternative to traditional clustering algorithms and serves as a useful tool for clustering tasks due to its flexibility and capacity to optimize parameters. There is room for further exploration and implementation of the Black-winged Hawk Optimization Algorithm in clustering techniques and practical data analysis problems for improvement.

Nowadays, it is very common to simulate animal and bird intelligence and behaviours to solve search and optimization problems. The Black Kite Optimization Algorithm (BKA), inspired by the migration and predatory habits of the black kite, is a metaheuristic optimization algorithm. The Black Kite optimization algorithm has been developed in this paper to solve a common clustering problem. The data is collected using the principle of maximum dissimilarity between data from different clusters and high similarity among data within the same cluster, and is grouped into clusters. The output of this algorithm contradicts the well-known TSA clustering methodology. The Black Kite optimization algorithm can be used to solve clustering problems based on pre-computation experience related to intra-cluster distance function and standard deviation.

Furthermore, the findings demonstrate the effectiveness, simplicity of implementation, and robustness of the proposed algorithm compared to TSA methodologies. Potential ways to enhance the performance of the proposed algorithm have also been identified. Exploring the integration of the BKA clustering algorithm with alternative clustering methods and using different fitness functions for clustering could be future research areas.

References

1. Boroujeni, S.P.H. and E. Pashaei. *Data clustering using chimp optimization algorithm*. in *2021 11th international conference on computer engineering and knowledge (ICCKE)*. 2021. IEEE.
2. Likas, A., N. Vlassis, and J.J. Verbeek, *The global k-means clustering algorithm.* Pattern recognition, 2003. 36(2): p. 451-461.
3. Rasooli, A.Q. and O. Inan, *Clustering with the Blackwinged Kite Algorithm.* International Journal of Computer Science & Communications (IJCSC), 2024. 9(1): p. 22-33.
4. Kao, Y. and K. Cheng. *An ACO-based clustering algorithm*. in *International Workshop on Ant Colony Optimization and Swarm Intelligence*. 2006. Springer.
5. Mirjalili, S., S.M. Mirjalili, and A. Lewis, *Grey wolf optimizer.* Advances in engineering software, 2014. 69: p. 46-61.
6. Lei, K., et al., *An NDN IoT content distribution model with network coding enhanced forwarding strategy for 5G.* IEEE Transactions on Industrial Informatics, 2017. 14(6): p. 2725-2735.
7. Abdullahi, M., et al., *A survey of symbiotic organisms search algorithms and applications.* Neural computing and applications, 2020. 32(2): p. 547-566.
8. Nasiri, J. and F.M. Khiyabani, *A whale optimization algorithm (WOA) approach for clustering.* Cogent Mathematics & Statistics, 2018. 5(1): p. 1483565.
9. Hatamlou, A., *Black hole: A new heuristic optimization approach for data clustering.* Information sciences, 2013. 222: p. 175-184.
10. Wang, G.-G., A.H. Gandomi, and A.H. Alavi, *Stud krill herd algorithm.* Neurocomputing, 2014. 128: p. 363-370.
11. Abualigah, L.M., A.T. Khader, and E.S. Hanandeh, *Hybrid clustering analysis using improved krill herd algorithm.* Applied Intelligence, 2018. 48(11): p. 4047-4071.
12. Mirjalili, S. and A. Lewis, *The whale optimization algorithm.* Advances in engineering software, 2016. 95: p. 51-67.
13. Wadhwa, A., S. Garg, and M.K. Thakur. *Automatic detection of DBSCAN parameters using BAT algorithm*. in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*. 2023.
14. Chen, Y.-R., M.-C. Hung, and D.-L. Yang, *Using data mining to construct an intelligent web search system.* International Journal of Computer Processing of Oriental Languages, 2003. 16(02): p. 143-170.
15. Raya-Tapia, A.Y., et al., *Fundamentals of Clustering: Methods, Metrics, and Optimization*, in *Machine Learning and Clustering for a Sustainable Future: Applications in Engineering and Environmental Science*. 2025, Springer. p. 13-50.
16. Sadeghi, V. and H. Etemadfard, *Optimal cluster number determination of FCM for unsupervised change detection in remote sensing images.* Earth Science Informatics, 2022. 15(2): p. 1045-1057.

Vol. 9, No. 1, 2024, pp. 22-33

http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

17. Özdemir, Ö. and A. Kaya, *Comparison of FCM, PCM, FPCM and PFCM algorithms in clustering methods.* 2019.
18. Xue, R., et al. *Multi-strategy Integration Model Based on Black-Winged Kite Algorithm and Artificial Rabbit Optimization*. in *International Conference on Swarm Intelligence*. 2024. Springer.