

CPSO: Chaotic Particle Swarm Optimization for Cluster Analysis

Jiaji Wang

School of Math and Information Technology, Jiangsu Second Normal University, Nanjing, Jiangsu 210016, People's Republic of China

(Received 15 February 2023; Revised 13 March 2023; Accepted 13 March 2023; Published online 14 March 2023)

Abstract: *Background:* To solve the cluster analysis better, we propose a new method based on the chaotic particle swarm optimization (CPSO) algorithm. *Methods:* In order to enhance the performance in clustering, we propose a novel method based on CPSO. We first evaluate the clustering performance of this model using the variance ratio criterion (VRC) as the evaluation metric. The effectiveness of the CPSO algorithm is compared with that of the traditional particle swarm optimization (PSO) algorithm. The CPSO aims to improve the VRC value while avoiding local optimal solutions. The simulated dataset is set at three levels of overlapping: non-overlapping, partial overlapping, and severe overlapping. Finally, we compare CPSO with two other methods. *Results:* By observing the comparative results, our proposed CPSO method performs outstandingly. In the conditions of non-overlapping, partial overlapping, and severe overlapping, our method has the best VRC values of 1683.2, 620.5, and 275.6, respectively. The mean VRC values in these three cases are 1683.2, 617.8, and 222.6. *Conclusion:* The CPSO performed better than other methods for cluster analysis problems. CPSO is effective for cluster analysis.

Keywords: cluster analysis; chaotic particle swarm optimization; variance ratio criterion

I. INTRODUCTION

Machine learning (ML) is one of the subfields of artificial intelligence (AI). Ref. [1] allows computers to explore ways to improve problem-solving efficiency based on experience [2,3]. Deep learning (DL) is a research direction in the field of ML [4,5]. Based on the learning method of the model, ML can be broadly classified into supervised, unsupervised, and semisupervised learning [6]. Clustering analysis is one of the unsupervised learning methods [7]. In cluster analysis, for a large amount of data or samples, the computer tries to find the classification rules based on the characteristics of the data. It follows this classification rule to classify the data reasonably. After several iterations, finally, similar data are classified into one group.

Cluster analysis applies to many different fields. Fahad, S. et al. (2023) [8] used cluster analysis to predict rainfall. Dreyer, J. et al. (2022) [9] looked for the common features and differences in the care that people with dementia currently receive in order to design the care that different levels of patients need. Azzi, D. V. et al. (2022) [10] used cluster analysis to determine students' quality of life by combining multiple factors to help teachers guide students in finding response strategies. Negi, S. S. et al. (2022) [11] used clustering and network analysis of over one million SARS-CoV-2 spike protein sequences to help doctors design vaccines for multiple local strains. Naciti, V. et al. (2022) [12] analyzed a large amount of literature on corporate governance and sustainability, finding trends recent in scholarly research in this area. In addition to the applications of cluster analysis in the above fields, it can also be used in agriculture [13], Internet-of-Things [14], and investment [15].

In recent years, numerous researchers have attempted to introduce new clustering algorithms from diverse angles to improve optimization. Liang, B. et al. (2022) [16] simulated the process of cell division and merging in living organisms and developed a novel two-stage algorithm for clustering cell groups. This algorithm involves a continuous validation and correction mechanism. This algorithm's advantage was that it applies to many dataset types, including imbalanced datasets and spherical nonspherical clusters. The paper [17] proposed a novel clustering algorithm. It started with a fast mean-shift algorithm to locate the center of the data distribution quickly. Next, the parameters in the algorithm were adjusted by calculating the distance distribution between the centroids and the other points. The process of this new clustering algorithm was similar to the radar scan. The aim is to classify a cluster efficiently and accurately.

The paper [18] presented the TPLMM-k algorithm, which pioneered the use of the asymmetric distribution of the k-means algorithm to recognize the existence of bacteria on digital images. Compared to the gaussian mixture model (GMM) method, which is currently used more frequently, better image decomposition metrics can be obtained using this novel method. The paper [19] proposed a metaheuristic clustering task algorithm based on water wave optimization (WWO). The original WWO method may have problems of missing optimal information and premature convergence. The improved method uses the updated search mechanism and the decay operator to solve these two problems separately.

The paper [20] wanted to classify customer types dynamically. The innovative framework presented in the study utilizes a time series of recency, frequency, and monetary variables to represent each customer's behavior. Experimental results showed that finding a suitable model for target customer groups in different industries is feasible. The framework proposed in the paper [21]

Corresponding author: Jiaji Wang (e-mail: wangjiaji10@gmail.com).

consisted of kernel fuzzy C-means clustering (KFCM) and a teaching learning-based optimization algorithm (TLBO). The strength was that it improved clustering compactness. Using the same dataset, the optimized clustering algorithm performs better than the genetic algorithm and the particle swarm algorithm for the model. The paper [22] developed a recommendation system for recommending content from users with similar hobbies to the new users. The hybrid sparrow clustered (HSC) system consists of two phases: training and recommendation. Clustering analysis was mainly used in the first phase to classify clusters based on a rating matrix generated from the data.

The paper [23] introduced the firefly algorithm (FA) to find the maximal point of the variance ratio criterion (VRC), which is the fitness function in this study. Wang, S. (2016) [24] proposed chaotic biogeography-based optimization (CBBO) and applied it to centroid-based clustering methods. Chandrasekaran, D. et al. (2019) [25] applied the cat swarm algorithm (CSA) to minimize the intra-cluster distances between the cluster members. The CSA method can be transferred to cluster analysis and bankruptcy prediction [26].

In the following part, Section II reviews related works and some classical clustering algorithms. Section III introduces the encoding strategy used and the criteria function. Section IV describes the PSO and chaotic particle swarm optimization (CPSO) algorithms. Section V contains simulation data and results of this algorithm compared with other optimization algorithms through VRC values. Finally, Section VI summarizes our work and suggests possible future directions for the improvement of this work.

II. RELATED WORK

Before cluster analysis, the data outliers from the dataset need to be removed, reducing interference of invalid samples to the cluster analysis results. Cluster analysis is a way of exploring the characteristics and rules that exist in the data instead of relying on experience gained before. The classification will follow these characteristics and rules to achieve a reasonable result.

Cluster analysis does not only classify by one particular characteristic. Considering multiple characteristics is necessary to obtain comprehensive and accurate classification results. The cluster analysis results may output different conclusions depending on the cluster method used. The number of final clusters obtained

from cluster analysis may differ from researcher to researcher, even based on the same dataset.

Some people have focused on optimization algorithms for clustering analysis. In the paper [27], a new density peak clustering algorithm (NDPC) is proposed. It was built based on a combination of cosine similarity and exponential decay function. With the enhance by an improved Harris hawks optimization algorithm (FHHO), the FHHO-NDPC algorithm has significant advantages in terms of convergence speed and solution precision when solving high-dimensional problems. A adaptive semisupervised clustering (A-SSC) algorithm was introduced in the paper [28]. This model was a combination of the nonsmooth optimization method and an incremental approach without the requirements of subgradient evaluations. By observing the circulation and convergence of water in the hydrologic cycle process, hydrologic cycle optimization (HCO) was proposed [29]. The HCO was abstracted into a simplified model. Tests were performed to evaluate the optimization effectiveness of the HCO algorithm. In this paper [30], the authors used a combination of the whale optimization algorithm and tabu search (WOATS) for data clustering. Regardless of the dataset's size, WOATS was able to efficiently identify high-quality centers in a limited number of iterations. The effectiveness of WOATS was demonstrated by the significant differences observed in different outlier criteria. Fig. 1 shows the classification of clustering analysis and features are listed in Table I.

III. METHODOLOGY

A. SEARCH SPACE: ENCODING STRATEGY

The dimension n of the search space is equal to the number of objects to be clustered and analyzed. k denotes the number of clusters hoped to be classified. j denotes the j th object. i denotes the i th individual. In the figure, $X_i = \{X_{i1}, X_{i2}, X_{i3} \dots X_{ij}\}, j \in [1, 12]$. The number of target clusters is $k = 3$. The first cluster contains the 1st, 4th, 7th, and 10th object. The second cluster contains the 2nd, 5th, 8th, and 11th object. The third cluster contains the 3rd, 6th, 9th, and 12th object. Fig. 2 shows the representation of encoding.

B. CRITERION FUNCTION

The VRC is a statistical test to determine the optimal number of clusters to a dataset. VRC measures the ratio of between-cluster

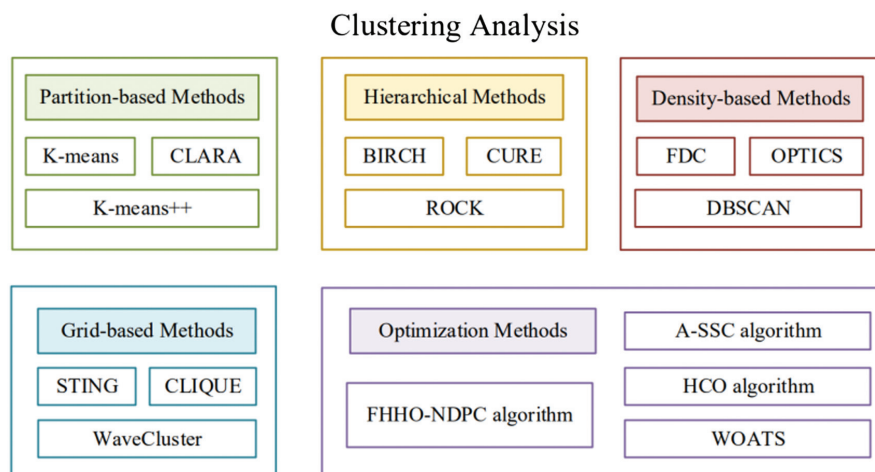
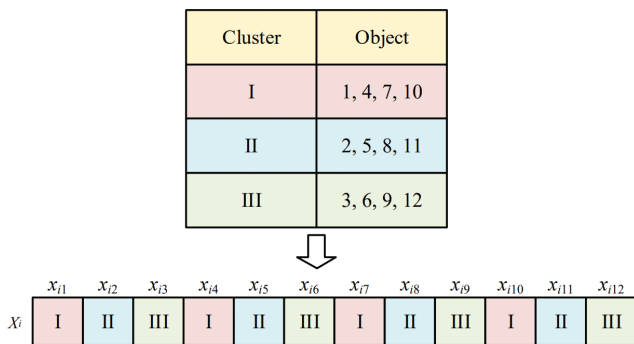


Fig. 1. Classification of clustering analysis.

Table I. Methods of clustering analysis

Type	Name	Feature
Partition-based methods	K-means [37]	Using the center of a cluster to represent a cluster
	K-means++ [38]	Improvements have been made in the way K-means performs the initialization of cluster centers
	CLARA [39]	Sampling techniques are used to enable the processing of large-scale data
Hierarchical methods	BIRCH [40]	Using a tree structure to process the dataset
	CURE [41]	Partitioning the dataset by randomly selected samples and global clustering after local clustering of each partition
	ROCK [42]	Taking the influence of surrounding objects into account when calculating the similarity of two objects
Density-based methods	DBSCAN [43]	Using spatial indexing techniques to search the neighborhood of an object, introducing the concept of density reachability
	OPTICS [44]	Different parameters can be set for different clusters
	FDC [45]	Dividing the entire data space into rectangular spaces
Grid-based methods	STING [46]	Using grid cells to preserve data statistics, thus enabling multi-resolution clustering
	Wave Cluster [47]	Introducing the wavelet transform, mainly used in signal processing
	CLIQUE [48]	Combining grid and density clustering algorithms
Others	Quantum Clustering [49]	Introducing the theory of quantum potential energy from quantum mechanics
	Kernel Clustering [50]	Mapping of kernel functions to highlight features not originally shown
	Spectral Clustering [51]	Clustering on arbitrarily shaped sample spaces and convergence to a globally optimal solution
Optimization methods	FHHO-NDPC algorithm [27]	Significant improvements in convergence speed and solution accuracy when solving high-dimensional problems
	A-SSC Algorithm [28]	Better than the other four algorithms in identifying compact and well-separated clusters
	HCO algorithm [29]	Abstracted from the hydrological cycling phenomenon
	WOATS [30]	Significant differences can be observed in different outlier criteria

**Fig. 2.** The representation of encoding.

variance to within-cluster variance. If the ratio is high, the data should be divided into more clusters; if the ratio is low, the data can be well represented by fewer clusters. The VRC requires balancing the separation between groups and the similarity within classes. VRC is a simple and effective method for determining the optimal number of clusters in a dataset. Therefore, the VRC is defined in equation, as follows:

$$\text{VRC}_k = \frac{\text{tr}(S_B)}{\text{tr}(S_W)} \times \frac{(N - k)}{(k - 1)}, \quad (1)$$

where k denotes the number of clusters, N denotes the number of objects in the dataset, B represents the between-group sum-of-squares cross-products matrix, and W represents the within-group sum-of-squares cross-products matrix. S_W denotes the average of the within-class scatter matrix, S_B denotes the between-class scatter matrix, and $\text{tr}()$ denotes the trace of the scattering matrices.

IV. OPTIMIZATION METHOD

In many algorithms, scholars often pay attention to the variation of the loss function. There is no doubt that we want the loss function to be as small as possible. Thus, it is necessary to use optimization methods. Commonly used optimization algorithms can be broadly classified into two categories, exact algorithms, and heuristics algorithms. Figure 3 shows the categorization of optimization algorithms.

Swarm intelligence optimization algorithms are inspired by the behaviors of birds flocking, ants searching for food, and bees searching for flowers. The optimization efficiency is improved by simulating the movement and information exchange of a large number of particles [31,32]. This method uses the knowledge and experience of each individual in the group to make collective decisions and realizes the search for the optimal solution through cooperation and collaboration within the group. This method is usually used for optimizing complex functions, such as finding the maximum or minimum value of a function [33] or combinatorial optimization [34].

A. PARTICLE SWARM OPTIMIZATION

The particle swarm optimization (PSO) algorithm is a commonly used algorithm in swarm intelligence optimization. The social behavior of birds inspires it in groups. The algorithm uses multiple particles to track the position of the best particle and the best position of itself that has been obtained so far.

Initially, the particles start from a random position and continuously transmit information to locate the optimal solution. In each information transfer, the particle updates the optimum

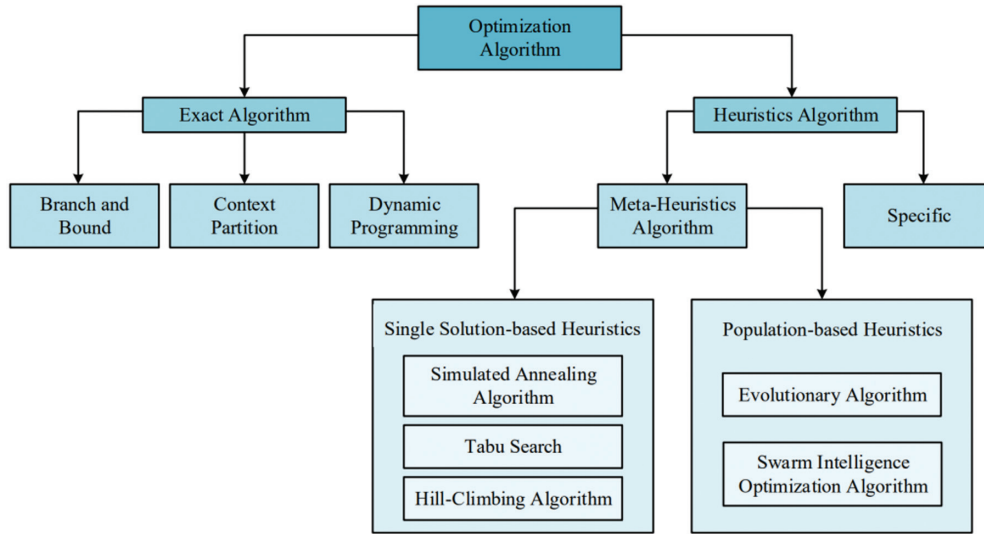


Fig. 3. Classification of optimization algorithms.

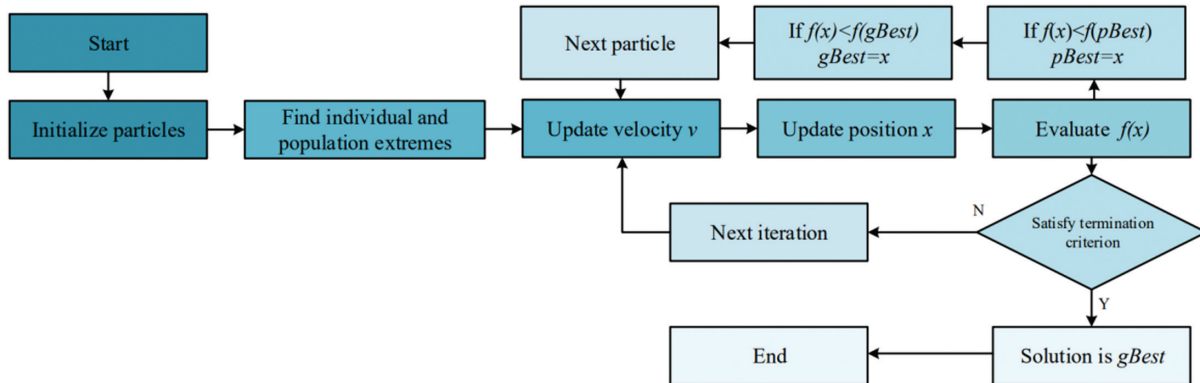


Fig. 4. The flow of the PSO algorithm.

solution it found and learns the optimum solution found by the whole group. The optimal solution discovered by a particle alone is referred to as the individual extremum, while the optimal solution currently found by the entire group is known as the global extreme. Through continuous iterations, the activities of the whole group in the space become continuously ordered to obtain the optimal solution eventually.

The PSO algorithm starts from a random solution and evaluates the quality of the solution through fitness during the iterative process. Compared with classical genetic algorithms, PSO is simpler to implement as there is no possibility of variation. The flow of the algorithm for finding the extremum according to the particle swarm algorithm function is shown in Fig. 4.

The formulas in the key steps of the PSO algorithm are the velocity vector iteration formula and the position vector iteration formula for particle i :

$$v_i = \omega v_i + c_1 r_1 (pBest_i - x_i) + c_2 r_2 (gBest - x_i) \quad (2)$$

$$x_i = x_i + v_i \quad (3)$$

Equation (2) is the iterative formula for the velocity vector v_i . Equation (3) is the iterative formula for the position x_i .

The parameter ω is called the inertia weight, which takes values in the interval $[0,1]$. The parameters c_1 and c_2 are learning factors. r_1 and r_2 are random probability values between $[0,1]$. $pBest_i$ represents the historical best position vector of particle i and the population's historical best position vector. $gBest$ represents the population's historical best position vector.

B. CHAOTIC PSO

In the chaos, even the smallest gaps that appear at the beginning can be infinitely magnified, leading to processes that cannot be predicted with different outcomes. Although chaos and randomness can make events unpredictable, the reasons are different. Chaos has no randomness. Randomness means that even the exact same initial conditions will measure different outcomes. The result is independent of the initial conditions.

Chaos leads to different results because of the extreme sensitivity of the results to the initial conditions. Chaotic optimization algorithms achieve global optima according to their laws of motion. It is easy to get rid of local optima and have better global optimization performance, which can compensate for the original PSO algorithm's tendency to fall into local extremes and slow convergence at a later stage. Figure 5 shows the flowchart of

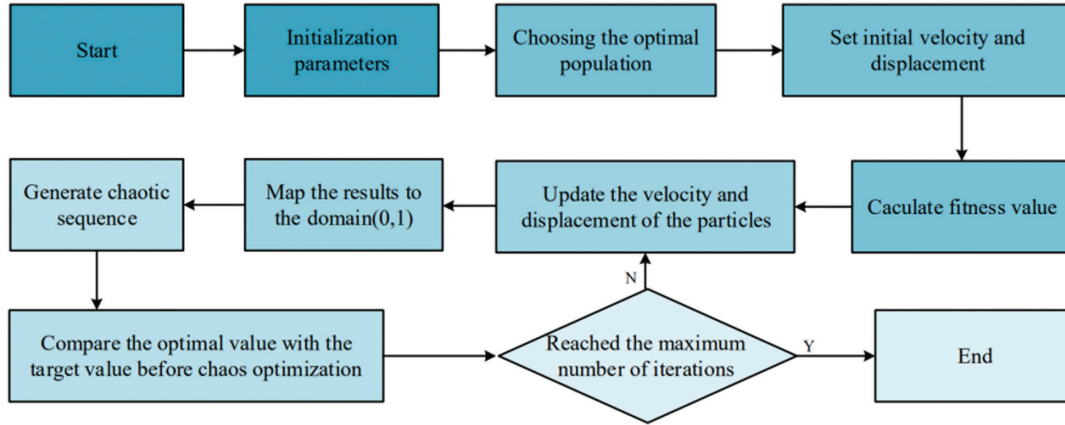


Fig. 5. Flowchart of chaotic PSO algorithm.

chaotic PSO. Logistic mapping is a typical chaotic system, and the mapping relationship is as follows:

$$z_{i+1} = \mu z_i(1 - z_i), \quad z_i \in (0,1], \quad (4)$$

where μ is the control variable. After the value of μ is determined, a deterministic time series $z_1, z_2, z_3 \dots$ can be iterated from any initial value $z \in [0,1]$. i represents the number of iterations.

V. EXPERIMENT RESULTS AND DISCUSSIONS

A. SIMULATION DATA

Assume $N = 400, k = 4$. The dimension of the feature vectors is 2. Undeniably, some data may belong to more than one class simultaneously. It is difficult to assign these data to a particular class with certainty. Overlapping means that an object can belong to more than one cluster simultaneously. We conducted experiments according to different degrees of overlapping. The distribution of the data for different cases of non-overlapping, partial overlapping, and severe overlapping is shown in Fig. 6. There are four types of data in Fig. 6, represented by blue triangles, purple inverted triangles, red crosses, and green diamonds. From Fig. 6(a)–6(c), the degree of mixing of different types of data is continuously

increasing. The images obtained by a good clustering method have dense samples within the same cluster and samples far enough apart between different clusters.

B. ALGORITHM COMPARISON

In this study, we compared the performance of the CPSO algorithm with the FA [23] and CSA [25] algorithms. Table II shows the VRC values of these algorithms under various conditions. When there was no overlapping, the VRC values of all three algorithms were the same, at 1683.2, whether it was the best VRC value, the mean VRC value, or the worst VRC value. The best VRC values were still the same when there was partial overlapping. However, in this condition, the FA algorithm achieved the highest mean VRC value of 618.2. The mean VRC values obtained by the three algorithms were not significantly different. The CPSO algorithm had the highest worst VRC value at 583.7. For severe overlapping conditions, the FA, CSA, and CPSO algorithms achieved the highest VRC value of 275.6, with mean VRC value of 221.5, 224.0, and 222.6, respectively, and worst VRCs of 133.9, 141.8, and 144.0, respectively. Overall, the CPSO algorithm obtained the highest VRC values for different types of overlapping conditions. CPSO's superior performance in the most severe overlapping condition compared to the other algorithms tested in the study is an important advantage. Since overlapping data are a common challenge in many practical applications, this advantage suggests that CPSO

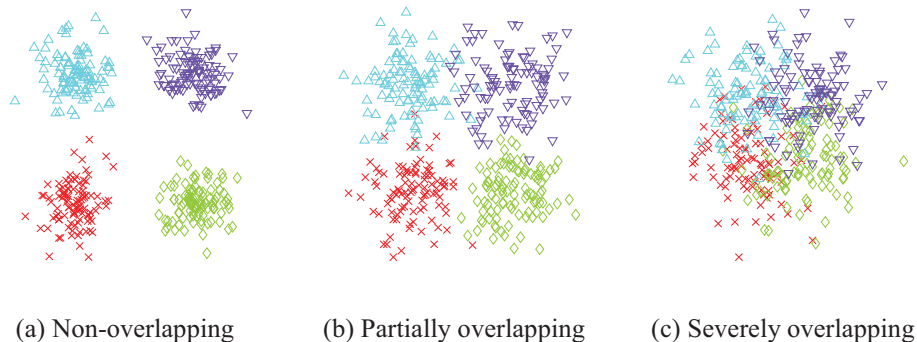


Fig. 6. Artificial data [24] to three different degrees.

Table II. Experiment results for three algorithms tested

Overlapping degree	VRC ↑	FA [23]	CSA [25]	CPSO (ours)
Non	Best	1683.2	1683.2	1683.2
	Mean	1683.2	1683.2	1683.2
	Worst	1683.2	1683.2	1683.2
Partially	Best	620.5	620.5	620.5
	Mean	618.2	616.3	617.8
	Worst	573.3	582.0	583.7
Severely	Best	275.6	275.6	275.6
	Mean	221.5	224.0	222.6
	Worst	133.9	141.8	144.0

(↑ means the larger, the better).

could be a valuable tool for handling real-world data that often contain varying degrees of overlap.

VI. CONCLUSION

An optimization algorithm based on CPSO for clustering analysis has been proposed in this paper, which has solved the limitations of traditional PSO in clustering analysis. The algorithm has introduced chaos factors into the PSO algorithm, which has prevented falling into local optimal solutions while ensuring global search. Comparing the VRC values of our proposed method with other methods, it has been found that our proposed method has better performance in different lapping conditions. In the future, we will improve the VRC values of this promising method in overlapping data. Also, many other optimization algorithms [36] will be tested.

References

- [1] X. Han, "A survey on deep learning in COVID-19 diagnosis," *J. Imaging*, vol. 9, no. 1, p. 1, 2023.
- [2] M. A. Makroum et al., "Machine learning and smart devices for diabetes management: Systematic review," *Sensors*, vol. 22, no. 5, p. 1843, 2022.
- [3] S. L. Clarke et al., "Future of machine learning in paediatrics," *Arch. Dis. Child.*, vol. 107, no. 3, pp. 223–228, 2022.
- [4] F. Saleem et al., "Machine learning, deep learning, and mathematical models to analyze forecasting and epidemiology of COVID-19: a systematic literature review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, p. 5099, 2022.
- [5] V. R. Allugunti, "Breast cancer detection based on thermographic images using machine learning and deep learning algorithms," *Int. J. Eng. Comput. Sci.*, vol. 4, no. 1, pp. 49–56, 2022.
- [6] D. Sharma et al., "Perceptron in supervised, semi-supervised, unsupervised learning and artificial neural network," *Central N.A. J. Theoretical Appl. Sci.*, vol. 3, no. 5, pp. 176–199, 2022.
- [7] C. Thongprayoon et al., "Distinct phenotypes of hospitalized patients with hyperkalemia by machine learning consensus clustering and associated mortality risks," *QJM*, vol. 115, no. 7, pp. 442–449, 2022.
- [8] S. Fahad et al., "Implementing a novel deep learning technique for rainfall forecasting via climatic variables: an approach via hierarchical clustering analysis," *Sci. Total Environ.*, vol. 854, p. 158760, 2023.
- [9] J. Dreyer et al., "Differences and commonalities of home-based care arrangements for persons living with dementia in Germany—a theory-driven development of types using multiple correspondence analysis and hierarchical cluster analysis," *BMC Geriatr.*, vol. 22, no. 1, pp. 1–21, 2022.
- [10] D. V. Azzi et al., "Quality of life, physical activity and burnout syndrome during online learning period in Brazilian university students during the COVID-19 pandemic: a cluster analysis," *Psychol Health Med.*, vol. 27, no. 2, pp. 466–480, 2022.
- [11] S. S. Negi et al., "Regional and temporal coordinated mutation patterns in SARS-CoV-2 spike protein revealed by a clustering and network analysis," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022.
- [12] V. Naciti et al., "Corporate governance and sustainability: a review of the existing literature," *J. Manage. Govern.*, vol. 26, no. 1, pp. 55–74, 2022.
- [13] N. D. DeLay et al., "Precision agriculture technology adoption and technical efficiency," *J. Agric. Econ.*, vol. 73, no. 1, pp. 195–219, 2022.
- [14] S. Mon et al., "Trust model for IoT using cluster analysis: a centralized approach," *Wirel. Pers. Commun.*, vol. 127, no. 1, pp. 715–736, 2022.
- [15] F. Soleymani et al., "Efficient portfolio construction by means of CVaR and k-means++ clustering analysis: evidence from the NYSE," *Int. J. Finance Econ.*, vol. 27, no. 3, pp. 3679–3693, 2022.
- [16] B. Liang et al., "A new cell group clustering algorithm based on validation & correction mechanism," *Expert Syst. Appl.*, vol. 193, p. 116410, 2022.
- [17] L. Ma et al., "A new clustering algorithm based on a radar scanning strategy with applications to machine learning data," *Expert Syst. Appl.*, vol. 191, p. 116143, 2022.
- [18] K. Satyanarayana et al., "Identifying the presence of bacteria on digital images by using asymmetric distribution with k-means clustering algorithm," *Multidimens. Syst. Signal Process.*, vol. 33, no. 2, pp. 301–326, 2022.
- [19] A. Kaur et al., "A new metaheuristic algorithm based on water wave optimization for data clustering," *Evol. Intell.*, vol. 15, 1, pp. 759–783, 2022.
- [20] H. Abbasimehr et al., "An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation," *Expert Syst. Appl.*, vol. 192, p. 116373, 2022.
- [21] S. Singh et al., "Kernel fuzzy C-means clustering with teaching learning based optimization algorithm (TLBO-KFCM)," *J. Intell. Fuzzy Syst.*, vol. 42, no. 2, pp. 1051–1059, 2022.
- [22] B. Sharma et al., "Hybrid Sparrow Clustered (HSC) algorithm for Top-N recommendation system," *Symmetry*, vol. 14, no. 4, p. 793, 2022.
- [23] Y. Zhang et al., "Cluster analysis by variance ratio criterion and firefly algorithm," *JDCTA: Int. J. Digital Content Technol. Appl.*, vol. 7, no. 3, pp. 689–697, 2013.

- [24] S. Wang, "A new logistic map based chaotic biogeography-based optimization approach for cluster analysis," in *Int. Conf. Prog. Inf. Comput. (PIC)*, Shanghai, China, IEEE, 2016, pp. 88–92.
- [25] D. Chandirasekaran et al., "Cat swarm algorithm in wireless sensor networks for optimized cluster head selection: a real time approach," *Cluster Comput.*, vol. 22, no. 5, pp. 11351–11361, 2019.
- [26] Y.-D. Zhang, "Cat Swarm optimization applied to alcohol use disorder identification," *Multimedia Tools Appl.*, vol. 77, no. 17, pp. 22875–22896, 2018.
- [27] W. Y. Guo et al., "Harris hawks optimization algorithm based on elite fractional mutation for data clustering," *Appl. Intell.*, vol. 52, no. 10, pp. 11407–11433, 2022.
- [28] A. M. Bagirov et al., "Nonsmooth optimization-based model and algorithm for semisupervised clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 2021, pp. 1–14, 2021.
- [29] X. H. Yan et al., "An adaptive hydrologic cycle optimization algorithm for numerical optimization and data clustering," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 6123–6151, 2022.
- [30] K. K. A. Ghany et al., "A hybrid modified step Whale Optimization Algorithm with Tabu search for data clustering," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 3, pp. 832–839, 2022.
- [31] Y. Zhang, "Optimization algorithms and machine learning techniques in medical image analysis," *Math. Biosci. Eng.*, vol. 20, no. 3, pp. 5917–5920, 2023.
- [32] S.-H. Wang, "Diagnosis of COVID-19 by Wavelet Renyi entropy and three-segment biogeography-based optimization," *Int. J. Comput. Intelligence Syst.*, vol. 13, no. 1, pp. 1332–1344, 2020.
- [33] M. Goerigk et al., "Combinatorial optimization problems with balanced regret," *Discrete Appl. Math.*, vol. 328, pp. 40–59, 2023.
- [34] M. Kumar et al., "Learning MAX-SAT from contextual examples for combinatorial optimisation," *Artif. Intell.*, vol. 314, p. 103794, 2023, Article ID. 103794.
- [36] S. Wang, "Pathological brain detection via Wavelet packet Tsallis entropy and real-coded biogeography-based optimization," *Fundam. Inform.*, vol. 151, no. 1–4, pp. 275–291, 2017.
- [37] F. Wu et al., "Research on image text recognition based on canny edge detection algorithm and k-means algorithm," *Int. J. Syst. Assur. Eng. Manage.*, vol. 13, no. 1, pp. 72–80, 2022.
- [38] N. Nugroho et al., "Penggunaan Metode K-means dan K-means++ Sebagai clustering data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170–179, 2022.
- [39] S. Gonzalo et al., "CLARA: a novel clustering-based resource-allocation mechanism for exploiting low-availability complementarities of voluntarily contributed nodes," *Future Gener. Comput. Syst.*, vol. 128, pp. 248–264, 2022.
- [40] H. Geranian et al., "Mapping of regional-scale multi-element geochemical anomalies using hierarchical clustering algorithms," *Nat. Resour. Res.*, vol. 31, no. 4, pp. 1841–1865, 2022.
- [41] J. Shi et al., "A novel hierarchical clustering algorithm with merging strategy based on shared subordinates," *Appl. Intell.*, vol. 52, no. 8, pp. 8635–8650, 2022.
- [42] A. T. Rumiati et al., "Mapping model for target achievement of village SDGs using the ensemble ROCK method: a case study of Sidoarjo Regency, East Java," *Int. J. Res. Bus. Soc. Sci. (2147–4478)*, vol. 11, no. 1, pp. 316–327, 2022.
- [43] E. Bajal et al., "A review of clustering algorithms: comparison of DBSCAN and K-mean with oversampling and t-SNE," *Recent Pat. Eng.*, vol. 16, no. 2, pp. 17–31, 2022.
- [44] Y. Li et al., "Visual analysis of vessel behaviour based on trajectory data: a case study of the Yangtze River estuary," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 4, p. 244, 2022.
- [45] N. Gokilavani et al., "Novel Fuzzy based density based clustering algorithm for effective cluster prioritization in WSN," *Int. J. Adv. Soft Comput. Appl.*, vol. 13, no. 2, pp. 123–138, 2021.
- [46] X. Zhu et al., "Nexus of mixed-use vitality, carbon emissions and sustainability of mixed-use rural communities: the case of Zhejiang," *J. Clean. Prod.*, vol. 330, p. 129766, 2022.
- [47] Y. Mao et al., "Comparison of wave-cluster and DBSCAN algorithms for landslide susceptibility assessment," *Environ. Earth Sci.*, vol. 80, no. 22, pp. 1–14, 2021.
- [48] F. Ren et al., "A novel high-dimensional trajectories construction network based on multi-clustering algorithm," *EURASIP J. Wirel. Commun. Netw.*, vol. 2022, no. 1, pp. 1–18, 2022.
- [49] O. Ouedrhiri et al., "Intelligent recommender system based on quantum clustering and matrix completion," *Concurr. Comput.: Pract. Exp.*, vol. 34, no. 15, p. e6943, 2022.
- [50] L. Li, "The intrusion data mining method for distributed network based on fuzzy kernel clustering algorithm," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 15, no. 1, pp. 32–45, 2022.
- [51] M. Ahmadi et al., "DQRE-SCnet: a novel hybrid approach for selecting users in federated learning with deep-Q-reinforcement learning based on spectral clustering," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7445–7458, 2022.