

A Wolf Pack Optimization Theory Based Improved Density Peaks Clustering Approach

Jinlong TIAN, Jianping ZHAO*, Wei ZHOU, Xuming HAN, Limin WANG*

Abstract: In view of the problem that the Density Peaks Clustering (DPC) algorithm needs to manually set the parameter cut-off distance (d_c) we propose a Wolf Pack optimization theory based Density Peaks Clustering approach (WPA-DPC). Firstly, we introduce d_c parameter into the Wolf Pack Algorithm (WPA) to speed up the search. Secondly, we introduce the WPA into the DPC algorithm; the cut-off distance is used as the location of the wolf group. Finally, we make silhouette index in the search process as the fitness value, and the optimal location of the wolf group is the parameter value at the end. The simulation results show that compared with the traditional Density Peaks Clustering algorithm, the proposed algorithm is closer to the true clustering number. According to the evaluation results of silhouette and f-measure, the quality of clustering and the accuracy are greatly improved.

Keywords: cut-off distance; density peaks clustering; evaluation index; wolf pack algorithm

1 INTRODUCTION

The traditional Density Peaks Clustering algorithm is determined by the selection of the value, and the clustering result is limited by the selection of the cut-off distance. In 2014, two famous scientists proposed a Density Peaks Clustering algorithm (DPC) [1]. This algorithm defines two kinds of information characteristics to express the data correlation [2]. The first is the relative density of the data points ρ , through the parameters d_c to measure the data points relative distance to other data points. The second one is the parameter δ of the data points, which is defined as the nearest distance between the data point and the high-density data point. The DPC believes that the chance of having comparatively high-density data points is far from either side. Then the decision matrix is established according to the characteristics of the information to locate the clustering center rapidly [3]. Subtle clever relationship structure design makes the algorithm quickly find the number of clusters, ignoring the cluster shape of the restrictions. And they can automatically mark the abnormal point. Density Peaks Clustering algorithm proposed in the Journal of Science has been well acknowledged in various fields recently.

Text clustering is one of the core problems in the field of text excavation and information search. The data similarity in class clusters is high and the data similarity between class clusters is small. The clustering algorithm can be divided into four categories: partitioned clustering algorithm, hierarchical clustering algorithm, clustering algorithm based on based and intelligent clustering algorithm. This paper mainly discusses the Density Peaks Clustering algorithm that was published by Alex Rodriguez and Alessandro Laio in Science in June 2014 [4]. Although the article has been questioned by many readers since it came out, the basic thinking of the new clustering algorithm is novel, simple and bright, which is worth learning. The core concept of this new clustering algorithm is to explain the clustering center. This paper will introduce the principle of this algorithm in detail and discuss some details.

Many years ago, people used particle swarm optimization (PSO) and ant colony optimization (ACO) for the use of ant colony, such as ants and birds, clustering intelligence algorithm, and successfully used to solve a

large number of practical problems [6-10]. The Wolf Pack Algorithm adopts bottom-up design method based on artificial wolves subject and collaborative search path structure based on responsibility division [5]. And individual behavior decision-making is based on their responsibilities [11]. Wolf detection: when searching for prey, wolves do not go out in full force, but send a few elite wolves to hunt within the possible range of the prey. The wolves around them would spontaneously run in the direction of the wolves and move closer to the prey [12, 13].

WPA is a new swarm intelligence algorithm put forward in 2013. This algorithm can be put into use to solve 0 - 1 knapsack problem, multidimensional knapsack problem, and optimal operation of hydropower station reservoir [14]. In this paper, Wolf Pack Algorithm is used to optimize the cut-off distance of Density Peaks Clustering algorithm, and the optimal results are found by adaptive algorithm [15]. This paper introduces the WPA to optimize the Density Peaks Clustering algorithm, so as to get better clustering results, improve clustering performance and make clustering more robust. The new method, which combines the clustering algorithm and swarm intelligence optimization algorithm, can process the financial data such as stocks and improve the clustering accuracy, which provides reliable evidence for fusion arithmetic. In part III, we introduced the Density Peaks Clustering algorithm based on the Wolf Pack Algorithm. In part IV, we made the experiment simulation and confirmed the advantages of novel algorithm.

2 DENSITY PEAKS CLUSTERING ALGORITHM

The DPC has been well acknowledged in various fields. In spite of this, DPC algorithm still has shortcomings. It is impossible to effectively process data points located in the low density region of the dataset, and mistakenly classify outliers and intermediate nodes into cluster classes. The artificial selection of cluster class center reduces the objectivity and accuracy of the algorithm to obtain the real cluster class. It cannot effectively process complex structure data, and performs poorly in processing complex data such as complex flow pattern, differentiation density and difference scale [16].

There are two significant regulations in the DPC algorithm. Cluster center point has a higher density than the other adjacent points. The cluster center point defined by the cluster has a higher density value point relatively far away [17, 18]. The original DPC algorithm defines the the distance between the data points by euclidean distance. Supposing that there are data points x_i, x_j , then the distance formula is as follows [19, 20].

$$d_{ij} = distance(x_i, x_j) \quad (1)$$

The algorithm process details are as follows: the clustering center is determined by the feature [21]; the other nodes are categorized into the nearest cluster with a higher density than themselves. DPC utilizes the distance between the cluster density node i and local-density of node j greater than i , and its definition is shown in Eq. (2) and (3).

$$\rho_i = \sum_j \chi \cdot (d_{ij} - d_c) \quad (2)$$

where d_{ij} is the distance about node i and j , d_c is the cut-off length. $\chi(x)$ is as follows:

$$\chi(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (3)$$

$$\delta_{qi} = \begin{cases} \min\{d_{qi, qj}\}, & i \geq 2 \\ \max\{\delta_{qi}\}, & i = 1 \end{cases} \quad (4)$$

The length with the highest local density δ_i is shown in Eq. (5):

$$\delta_i = \max(d_{ij}) \quad (5)$$

The parameter δ_i of the data node x_i which has the max local-density is named as the max distance between data point x_i and the others in the dataset. Otherwise, δ_i is defined as the nearest distance between the data node and the higher local-density node [22].

As shown in Eq. (2), the local-density ρ_i of the data nodes is influenced by the parameter d_c . Also, the algorithm shows that the experiment results of the DPC algorithm are less affected by d_c when the number of datasets is large. Otherwise, it would be severely affected by the d_c . The choice of the d_c affects the local density and clustering results of the data points. When the dataset sample is small, the DPC algorithm uses a Gaussian kernel function to calculate the sample density. The specific expression is shown in Eq. (5):

$$\rho_i = \sum_{j \neq i} \exp\left(-\left(\frac{d_{ij}}{d_c}\right)^2\right) \quad (6)$$

The local-density ρ_i and distance δ_i values of each data point are output according to the above formula, and then the corresponding decision graph can be generated to

determine and select the clustering center point. The clustering centroids often have large local density ρ_i and distance δ_i values at the same time, and based on this feature, the clustering centroids can be circled in the decision graph [23-25].

Table 1 The DPC algorithm

Input Distance matrix $\chi R_{N \times M}$, stage distance d_c	
Output Class label: $\gamma R_{N \times M}$	
Step 1	Calculate the cut-off distance d_c ; According to the Eq. (1), calculate the distance between the data points $d_{ij} R_{N \times M}$; Sort the data points in ascending order; Select the appropriate d_c according to the sorted distance matrix;
Step 2	Select the cluster center; Calculate ρ_i according to Eq. (2) and Eq. (3); Calculate δ_i according to Eq. (4); Sort the density of all alternative points in descending order; An alternative point with a high ρ value and a relatively high delta value is marked as the cluster center point;
Step 3	Assign the non clustered center point to the corresponding cluster center; According to the relation of δ_i value, the non-cluster center point is sequentially attached to its higher density point; Detecting presence of Halo points in the cluster.

3 DENSITY CLUSTERING ALGORITHM BASED ON WOLF PACK ALGORITHM

3.1 Wolf Pack Algorithm

Harsh living conditions and hundreds of years of evolution have created wolves with sophisticated organizational systems and sophisticated cooperative hunting methods. Wolves live in groups and have a clear social division of labor. They are responsible for the survival and development of wolves. In the initial solution object, the artificial wolf including the optimum value of the initial object function as the first-wolf. During iteration, the target function value of the optimum population after each iteration was compared with that of the previous generation of wolves. If the situation improves, it will update the location of opposing wolves. Instead of performing the three intelligent behaviors, the wolf simply proceeds to the other stage until it is changed by a more powerful artificial wolf.

At the first p ($p = 1, 2, \dots, h$) the location of exploratory wolves I in the d -dimensional situation after advancing in all directions.

$$x_{id}^{k+1} = x_{id}^k + step_b^d \cdot (\delta_d^k - x_{id}^k) / |\delta_d^k - x_{id}^k| \quad (7)$$

During the attack, suppose that the wolf perceives the prey odor concentration J is $x_j > X_{lead}$, and $X_{lead} = X_j$, the wolf transforms into a better wolf and initiates the evocation mechanism. Supposing $X_j < X_{lead}$, and the wolf J continues to attack till the length between it and the wolves is less than d_{near} , that is to turn to the siege. Supposing the variable d value to optimal is (minimum d , maximum d), then the determination distance d_{near} can be estimated by Eq. (8).

$$d_{near} = \frac{1}{D \cdot w} \cdot \sum_{d=1}^D |\min_d - \max_d| \quad (8)$$

The rushing wolves were near the prey, and the wolf had to join the scouts in a close siege to capture it. The

location of the wolves is nearest the prey, leader wolf, seen here as the location of the prey. Particularly, about the k -generation wolves, if the site of prey in the d -dimensional situation is set, then the siege action of wolves can be described by Eq. (9).

$$x_{id}^{k+1} = x_{id}^k + n \cdot step_b^d \cdot \left| \delta_d^k - x_{id}^k \right| \quad (9)$$

Supposing the data range of the variable d to be optimum (minimum d , maximum d), and $step_a$, $step_b$, $step_c$ and attack step in d -dimensional circumstances will be involved in three kinds of intelligent behaviors.

$$step_a^b = \frac{step_b^d}{2} = 2 \cdot step_c^d = \left| \max_d - \min_d \right| \quad (10)$$

Taking the maximum value of the solution function as an example, the process of solving Wolf Pack Algorithm is briefly described.

Step1: Spatial coordinates of wolves are randomly initialized in the solution space, and the artificial wolves are selected according to the size of the target function value.

Step 2: The wolves scout start to wander randomly to search for prey. If it finds that the target function value of a certain position is better than the object function value of the wolves, the position of the wolves will be updated. And the wolves will issue summoning behavior at the same time. If it is not found, the scout will continue to travel until it reaches its maximum number of rovers.

Step 3: The wolves that heard the call of the head wolf rushed to the head wolves with large step length. If the object function answer of the wolves were better than the object function answer of the head wolves during the running attack then the position of the opposite wolves would be updated. Otherwise, the wolves will continue to attack until they reach the siege area.

Step 4: Approaching the wolves will cooperate with the wolves scouts to round up the prey (the position of the wolves is regarded as the prey). If the target function value of other artificial wolves is better than that of the wolves, the position of the wolves will be updated until the prey is captured.

Step 5: The artificial wolves with small target function value in wolves were eliminated. And the new artificial wolves were randomly generated in the solution space to realize the renewal of wolves.

Step 6: Finally, determine whether wolf target function answer can meet the algorithm requirements, or whether all iterations in the process have been completed. If it does not reach the maximum number of iterations, then continue the iteration from the wolves detection walk until it reaches the maximum number of iterations. Otherwise, the output is the spatial coordinate and value of the head wolf objective function, which is the optimal solution of the function.

3.2 Density Peaks Clustering Algorithm Based on Wolf Pack Algorithm

By using the optimization effect of Wolf Pack Algorithm, the selection of cut-off distance d_c of Density Peaks Clustering algorithm is optimized. Density Peaks

Clustering algorithm clustering results are limited to the selection of cut-off distance. Wolf Pack Algorithm can be an index selection, optimization, to get the best value, the two algorithms to optimize, we can get better clustering results.

Table 2 Density Peaks Clustering based on Wolf Pack Algorithm

Input: Intercept distance d_c value, maximum iteration t , population size of wolves, location of wolves	
Output: Cluster number evaluation index Sil , Fm values	
Step 1	In the solution space, the spatial coordinates of wolves are randomly initialized, and the artificial wolves are selected according to the value of the objective function;
Step 2	If the target function value of a certain position is found to be greater than that of the target function value of the wolf, the position of the wolf will be renewed .If it is not found, the wolf scout will continue to travel until it reaches its maximum number of rovers;
Step 3	Run DPC algorithm and use internal index SIL to evaluate the effectiveness of each Wolf pack;
Step 4	Record the most effective Wolf position, iterate and output the clustering results, otherwise return to Step 2.

4 SIMULATION EXPERIMENT AND ANALYSIS

4.1 Evaluation Index

The experimental environment simulation was Inter (R) Pentium 2.9 GHz, memory for the 4.00 GB, hard drive 500 G, the operating system for the Window 7, programming language for the MATLAB 2019a.

For the good clustering results obtained in the experimental analysis and the known clustering number the clustering algorithm results were represented by N and compared with the previously known clustering algorithm results M . This way is named the external evaluation index analysis. We have the four relationships between m and n . This is a clustering result of a known structure advance evaluation method based on M .

(1) m and n pertain to the category of P , and also to the same partition in M .

(2) m and n are part of the category of P , but different from segmentation in M .

(3) m and n do not pertain to the different sample in P , but belong to the same segmentation in M .

(4) m and n are not part of the class in P , and are not part of the same segmentation in M .

4.1.1 Silhouette Indicators

Suppose there are N cluster classes in a dataset D containing n clusters: C_j ($i = 1, 2, \dots, n$). $a(t)$ represents the mean dissimilarity. $d(t, C_j)$ represents the mean dissimilarity or distance between all data points of C_j to another cluster C_j . $b(t) = \min\{d(t, C_j)\}$, where $j = 1, 2, \dots, n$ and $i \neq j$. The silhouette index calculation for sample t is shown in Eq. (11).

$$Sil(t) = \frac{(b(t) - a(t))}{\max\{a(t), b(t)\}} \quad (11)$$

The size of the $Sil(t)$ value reflects how good the clustering results are. It usually takes a value between -1 and 1 , and the closer the value is to 1 , the better the clustering result is.

4.1.2 F-measure Indicators

The *F*-measure indicator is an external indicator that unites the recall $R(i, j)$ and the accuracy rate $P(i, j)$. Supposing the true clustering P_j and C_i , the accuracy $P(i, j)$ and recall rate $R(i, j)$ are shown below:

$$P = \text{precision}(i, j) = \frac{N_{ij}}{N_i} \quad (12)$$

$$R = \text{recall}(i, j) = \frac{N_{ij}}{N_j} \quad (13)$$

F-measure index formula as shown in Eq. (14):

$$F(P_j, C_i) = \frac{2 \cdot P(P_j, C_i) \cdot R(P_j, C_i)}{P(P_j, C_i) + R(P_j, C_i)} \quad (14)$$

The magnitude of the *F*-measure value reflects the precision of the clustering result, and it takes a value between -1 and 1 . The closer the value is to 1 , the higher the clustering precision is.

$$F(i) = (\partial^2 + 1)PR / \partial^2(P + R) \quad (15)$$

In the simulation experiment, in order to verify the effectiveness of the experiment and the feasibility of WPA-DPC, after searching the data, this paper selects the Iris, Flame, Spiral, Seeds and Aggregation datasets of UCI dataset. The datasets used in this paper are shown in Tab. 3:

Table 3 UCI datasets

dataset	Number of samples	Dimension	Number of classes
Spiral	312	2	2
Flame	240	2	2
Aggregation	788	2	7
Iris	150	4	3
Seeds	210	7	3

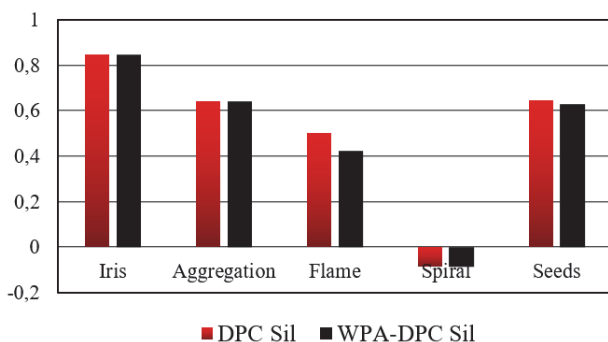


Figure 1 Comparison of Silhouette indicators

The WPA-DPC algorithm and DPC two algorithms are contrasted, mainly from two aspects to verify the advantages and disadvantages of the WPA-DPC algorithm so that we can determine the effectiveness with two kinds of clustering algorithm. The five datasets in UCI, DPC and WPA-DPC respectively. Then validate two algorithms accuracy rate (*FM*) and evaluate the clustering results index (*Sil*). We utilize the effectiveness of these indicators to

analyze the experimental results. The experimental results are shown in Fig. 1 and Fig. 2.

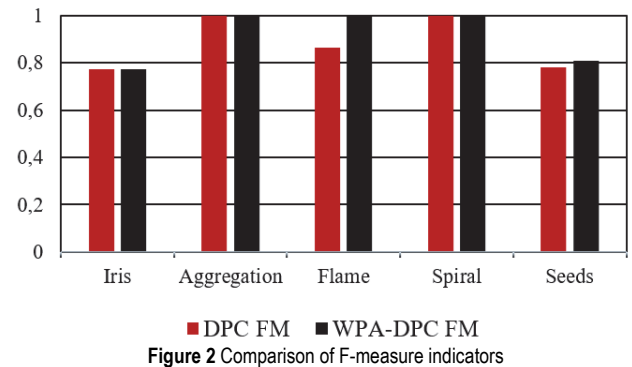


Figure 2 Comparison of F-measure indicators

Table 4 Comparison of Evaluation Indexes of Clustering performance

dataset	DP		WPA-DPC	
	FM	Sil	FM	Sil
Iris	0,7710	0,8462	0,7710	0,8462
Aggregation	1	0,6419	1	0,6419
Flame	0,8650	0,5007	1	0,4250
Spiral	1	-0,0867	1	-0,0867
Seeds	0,7803	0,6442	0,8092	0,6287

4.2 Cluster Performance and Experimental Results Analysis

It can be indicated from the *FM* index that the performance of the WPA-DPC algorithm and the traditional DPC algorithm in the five test datasets are generally equally those of the DPC algorithm. We can obviously see that the accuracy rate is improved. It can be seen from the silhouette index that the WPA value of the WPA-DPC algorithm in the dataset Iris, Aggregation Wine and Flame is equal to the traditional DPC algorithm. In the datasets Iris and Flame test comparison, the *Sil* index of the improved algorithm and the original DPC algorithm *sil* indicators are roughly the same. When the dataset is spherical data, the *sil* index is as high as the better. When the similar in the spiral structure of these complex structure. The *Sil* index and the quality of the cluster is a negative correlation between, so we can know WPA-DPC clustering better. In general, WPA-DPC's silhouette index is slightly higher than the DPC algorithm, the algorithm clustering results are better. In conclusion, the Wolf Pack Algorithm based on Density Peaks Clustering can effectively establish the similarity relation between data. It can find the optimal data value, the clustering situation is more evident, and the clustering result is more accurate.

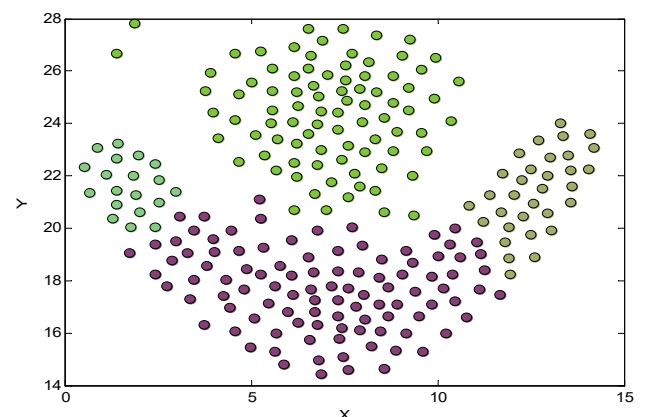


Figure 3 Flame dataset clustering result graph of DPC

From the comparison of the algorithm results above, we know that the clustering results of Flame datasets in DPC are similar to the results of WPA-DPC clustering. And their *sil* and *FM* index are the same, and each index data is very good. It shows that the clustering effect of WPA-DPC is also very good. To some extent, this improved clustering algorithm is very effective for Wolf Pack Algorithm based on DPC. Through the optimization of the algorithm parameters, the Wolf Pack Algorithm multiple operations, the final optimization parameter values are close to the same. The Result shows that Density Peaks Clustering algorithm based on WPA is excellent and sufficient. It has a positive effect on clustering results.

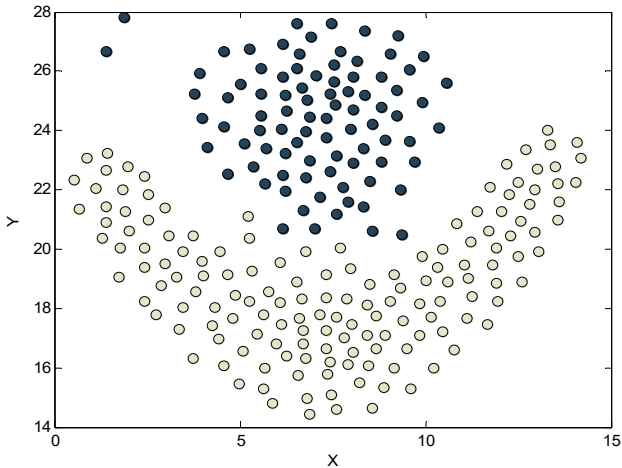


Figure 4 Flame dataset clustering result graph of WPA-DPC

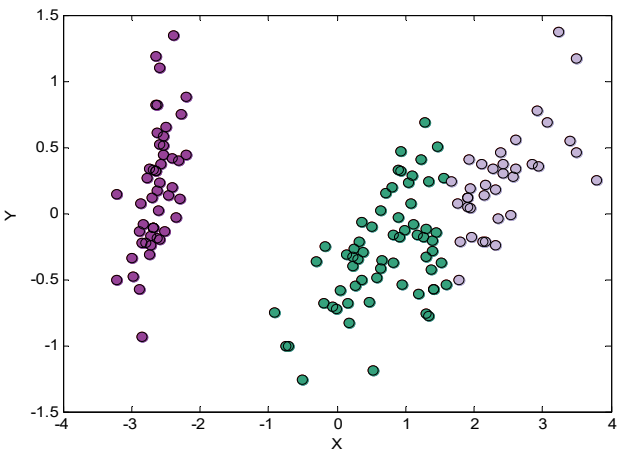


Figure 5 Iris dataset clustering result graph of DPC

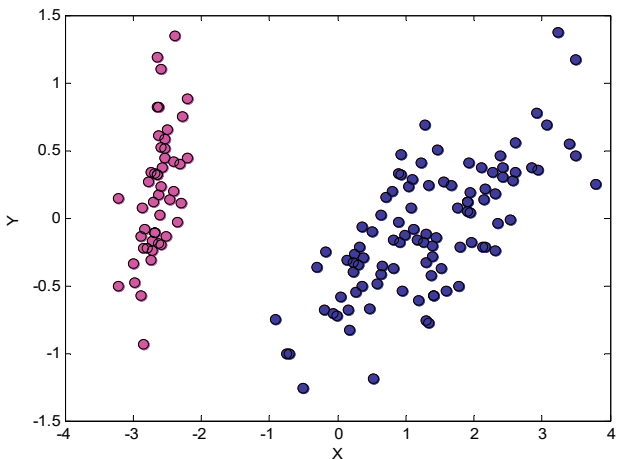


Figure 6 Iris dataset clustering result graph of WPA-DPC

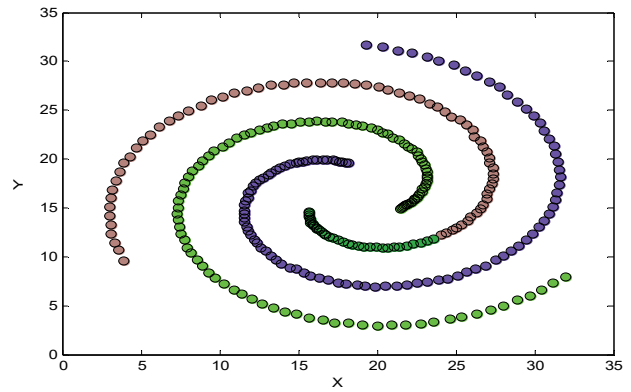


Figure 7 Spiral dataset clustering result graph of DPC

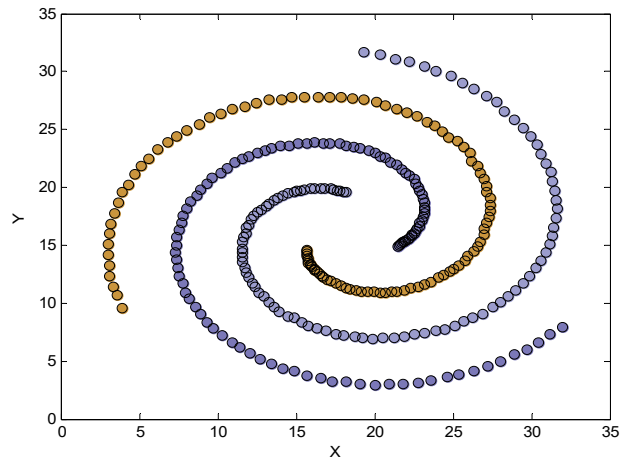


Figure 8 Spiral dataset clustering result graph of WPA-DPC

From the above figure, it can be concluded that the DPC cluster is divided into three categories and the distribution is scattered. And the distribution of WPA-DPC is obvious and concentrated. WPA-DPC makes the clustering result more impressive.

The Spiral dataset is spherical data, the better the *sil* index, the better the experiment results. The quality of clustering is negatively correlated, which indicates that the clustering results are better. The Wolf Pack Algorithm based on the Density Peaks Clustering can effectively establish the similarity connections between data, and find the optimal dataset value. The clustering trend is more obvious, and the clustering results are more accurate. The silhouette index of WPA-DPC is slightly equal to DPC algorithm, and the experimental results of this algorithm are better than those of previous ones.

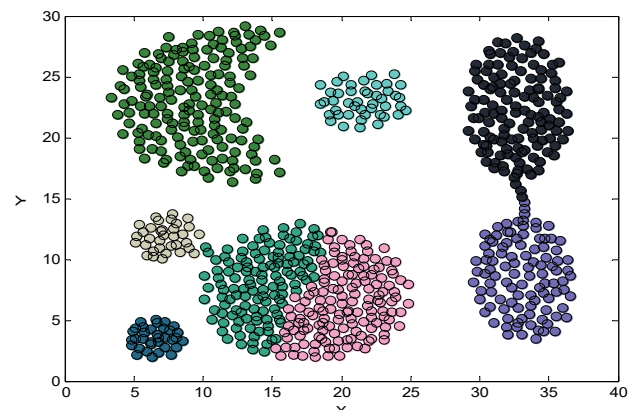


Figure 9 Aggregation dataset clustering result of DPC

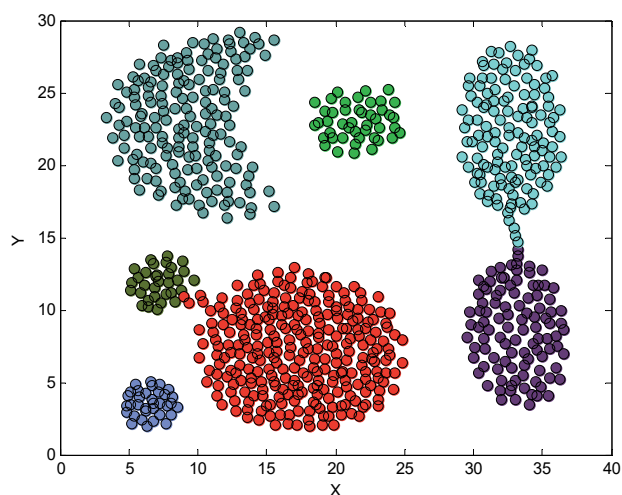


Figure 10 Aggregation dataset result of WPA-DPC

As shown in Fig. 8, it is the clustering result graph of aggregation of dataset, whose real class number is seven. It can be inferred from the clustering effects in the figure above that the improved algorithm can get the real number of classes, the number of clusters is accurate and the clustering efficiency is good.

It is very straightforward to see that the WPA-DPC algorithm is better than DPC algorithm. Therefore, the clustering algorithm based on coda-bird optimization algorithm (WPA-DPC) has a good propelling effect.

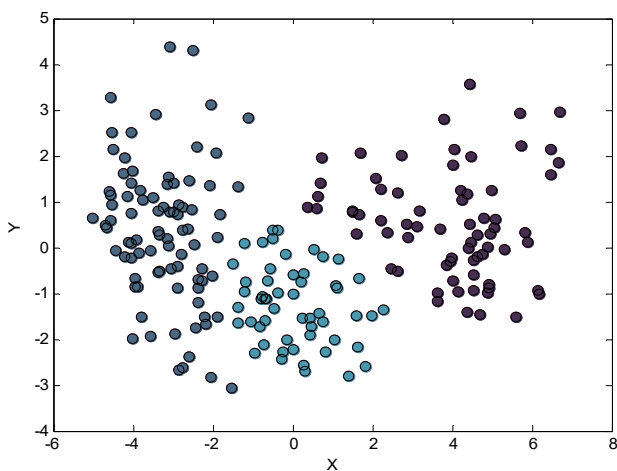


Figure 11 Seeds dataset clustering result graph of DPC

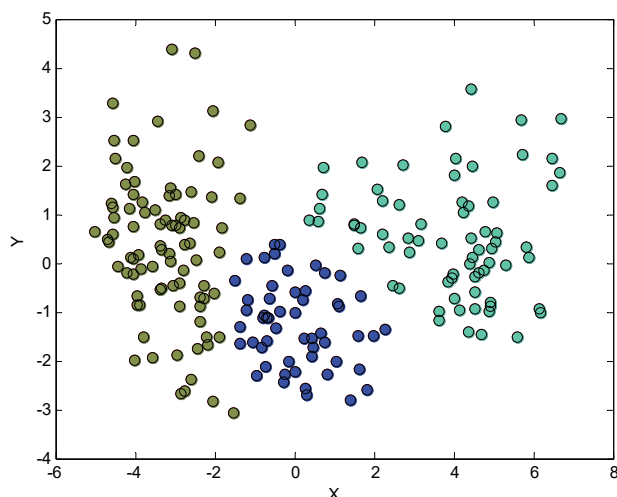


Figure 12 Seeds dataset clustering result of WPA-DPC

As shown in the figures, we get a clear clustering result diagram by optimizing the density Peaks clustering algorithm by Wolf Pack Algorithm. To verify the universal applicability of WPA-DPC, we utilize high dimensional dataset to verify the algorithm. Based on the simulation results, we can see that the method is also robust and effective to high-dimensional dataset.

The above three comparison diagrams show that the WPA-DPC algorithm is superior to the original DPC algorithm. The improved algorithm can automatically search for the optimal cut-off distance and achieve the best clustering result.

5 CONCLUSION

We proposed a Density Peaks Clustering algorithm (WPA-DPC) based on Wolf Pack Algorithm. The algorithm is combined with the idea of the optimization of the Wolf Pack Algorithm. The similarity matrix is established and the parameter value d_c of the DPC algorithm is updated. The results indicate that the clustering number of the improved algorithm is closer to the real number of clustering. The method performance and clustering evaluation index value are significantly improved.

At the same time, WPA-DPC algorithm can solve the problem of manually selecting parameters of DPC algorithm. However, the clustering effect is poor when dealing with complex datasets. Therefore, how to establish a link between different datasets and parameters is a problem that needs to be addressed. On the other hand, we optimize the different parameters, and then integrate all the optimization results to make the Density Peaks Clustering algorithm have better clustering effect. Whether these ideas can improve the quality of clustering remains to be studied. Our next step is to improve the DPC algorithm to adapt to the complex structure of the dataset.

Acknowledgements

This research is supported by the National Science Foundation of China under Grant No. 61202306, 61572225 and 61472049, the foundation of Jilin Province Education Department under Grant No. JJKH20190724KJ, the Jilin Provincial Science & Technology Department Foundation under Grant No. 20190302071GX and 20200201164JC, and the Development and Reform Commission Foundation of Jilin Province under Grant No. 2019C053-11.

6 REFERENCES

- [1] Han, X, Liu, Q, Wang, H, et al. (2018). Novel Fruit Fly Optimization Algorithm with Trend Search and Co-evolution. *Knowledge-Based Systems*, 141, 1-17. <https://doi.org/10.1016/j.knsys.2017.11.001>
- [2] Wang, L., Wu, L. L., & Fu, D. M. (2014). A density-based fuzzy adaptive clustering algorithm. *Beijing Keji Daxue Xuebao/Journal of University of Science & Technology Beijing*, 36(11), 1560-1566.
- [3] Mingjing, D., Shifei, D., & Hongjie, J. (2016). Study on Density Peaks clustering based on k-nearest neighbors and principal component analysis. *Knowledge-Based Systems*, 99, 135-145. <https://doi.org/10.1016/j.knsys.2016.02.001>

- [4] Rodriguez, A. & Laio, A. (2014). Clustering by fast search and find of Density. *Peakss.Science*, 344(6191), 1492-1496. <https://doi.org/10.1126/science.1242072>
- [5] Xu, J., Wang, G., & Deng, W. (2016). Den PEHC: Density Peaks based Efficient Hierarchical Clustering. *Information Sciences*, 373(12), 200-218. <https://doi.org/10.1016/j.ins.2016.08.086>
- [6] Gaidhane, P. J. & Nigam, M. J. (2018). A Hybrid Grey Wolf Optimizer and Artificial Bee Colony algorithm for Enhancing the Performance of Complex Systems. *Journal of Computational Science*, 27, 284-302. <https://doi.org/10.1016/j.jocs.2018.06.008>
- [7] Ji, C. & Lei, Y. (2017). Parallel clustering by fast search and find of density Peakss. *International Conference on Audio, Language and Image Processing, IEEE*, 563-567.
- [8] Juanying, X., Hongchao, G., & Weixin, X. (2016). K-nearest neighbor optimization for density Peaks fast search clustering algorithm. *Chinese science: information science*. (02), 258-280. <https://doi.org/10.1360/N112015-00135>
- [9] Wang, L., Tao, X., & Han, X. (2016). Semi-supervised affinity propagation based on Density Peakss semi-supervised affinity propagation based on Density Peakss. *Tehnički vjesnik*, 23(6), 1787-1794. <https://doi.org/10.17559/TV-20160531025240>
- [10] Mehmoood, R., Bie, R, Dawood, H., et al. (2015). Fuzzy Clustering by Fast Search and Find of Density Peakss. *2015 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI). IEEE*, 258-261. <https://doi.org/10.1109/IIKI.2015.62>
- [11] Wang, S., Wang, D., Li, C. et al. (2016). Clustering by Fast Search and Find of Density Peakss with Data Field. *Chinese Journal of Electronics*, 25(3), 397-402. <https://doi.org/10.1049/cje.2016.05.001>
- [12] Jianhua, J., Xing, T., Limin, W., et al. (2016). An Improved Recognition Method on Halo Node for Density Peaks Clustering Algorithm. *2016 The International Conference on Management of e-Commerce and e-Government*, 282-293.
- [13] Clara, P. & Domenico, T. (2003). P-Auto Class: scalable parallel clustering for mining large datasets. *IEEE Trans on Knowledge and Data Engineering*, 15(3), 629-641. <https://doi.org/10.1109/TKDE.2003.1198395>
- [14] Xu, X., Ding, S., Xu, H., et al. (2018). A feasible density Peakss clustering algorithm with a merging strategy. *Soft Computing*, 1-13. <https://doi.org/10.1007/s00500-018-3183-0>
- [15] Liu, W. & Hou, J. (2017). Study on a density Peaks based clustering algorithm. *International Conference on Intelligent Control & Information Processing, IEEE*, 60-67. <https://doi.org/10.1109/ICICIP.2016.7885877>
- [16] Xie, J., Gao, H. & Xie, W. (2016). K-nearest neighbors optimized clustering algorithm by fast search and ending the density Peakss of a dataset. *Scientia Sinica Informationis*, 46(2), 258-280. <https://doi.org/10.1360/N112015-00135>
- [17] Han, X. H., Quan, L., Xiong, X. Y., et al. (2017). A novel data clustering algorithm based on modified gravitational search algorithm. *Engineering Applications of Artificial Intelligence*, 61, 1-7. <https://doi.org/10.1016/j.engappai.2016.11.003>
- [18] Wang, J., Xia, L. N., & Jing, J. W. (2009). Maximum density clustering algorithm. *Journal of the Graduate School of the Chinese Academy of Sciences*, 26(4), 539-548.
- [19] Du, M., Ding, S., Xu, X., et al. (2017). Density Peakss clustering using geodesic distances. *International Journal of Machine Learning & Cybernetics*, 1-15. <https://doi.org/10.1007/s13042-017-0648-x>
- [20] Parmar, M., Wangž, D., Tan, A. H., et al. (2017). A novel density Peaks clustering algorithm based on squared residual error. *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). IEEE*, 348, 82-96. <https://doi.org/10.1109/SPAC.2017.8304248>
- [21] Kumar, Y. & Sahoo, G. (2014). A Review on Gravitational Search Algorithm and its Applications to Data Clustering & Classification. *International Journal of Intelligent Systems & Applications*, 6(6), 79-93. <https://doi.org/10.5815/ijisa.2014.06.09>
- [22] Nikbakht, H. & Mirvaziri, H. (2015). A new algorithm for data clustering based on gravitational search algorithm and genetic operators. *International Symposium on Artificial Intelligence and Signal Processing. IEEE*, 222-227. <https://doi.org/10.1109/AISP.2015.7123532>
- [23] Rong, Z., Yong, Z., Shengzhong, F., et al. (2018). A Novel Hierarchical Clustering Algorithm Based on Density Peakss for Complex Datasets. *Complexity*, 1-8. <https://doi.org/10.1155/2018/9489620>
- [24] Huang, L., Wang, G., Wang, Y., et al. (2016). A link density clustering algorithm based on automatically selecting density Peakss for overlapping community detection. *International Journal of Modern Physics B*, 30(24), 165-167. <https://doi.org/10.1142/S0217979216501678>
- [25] Burgan, H. I. & Aksoy, H. (2020). Monthly Flow Duration Curve Model for Ungauged River Basins. *Water*, 12, 338. <https://doi.org/10.3390/w12020338>

Contact information:

Jinlong TIAN, PhD Candidate,
School of Computer Science and Engineering,
Changchun University of Science and Technology, Changchun130022, China
E-mail: 346119380@qq.com

Jianping ZHAO, Professor
(Corresponding author)
School of Computer Science and Engineering,
Changchun University of Science and Technology, Changchun130022, China
E-mail: 775388410@qq.com

Wei ZHOU, PhD Candidate
School of Computer Science and Engineering,
Changchun University of Science and Technology, Changchun130022, China
E-mail: 2020200125@mails.cust.edu.cn

Xuming HAN, Professor
College of Information Science Technology, Jinan University, Guangzhou,
510632, China
E-mail: hanxuming@163.com

Limin WANG, Professor
(Corresponding author)
School of Internet Finance and Information Engineering, Guangdong University
of Finance, Guangzhou, 510521, China
E-mail: wlm_new@163.com