

IMPROVED FRUIT FLY OPTIMIZATION ALGORITHM-BASED DENSITY PEAK CLUSTERING AND ITS APPLICATIONS

Ruihong Zhou, Qiaoming Liu, Zhengliang Xu, Limin Wang, Xuming Han

Original scientific paper

As density-based algorithm, Density Peak Clustering (DPC) algorithm has superiority of clustering by finding the density peaks. But the cut-off distance and clustering centres had to be set at random, which would influence clustering outcomes. Fruit flies find the best food by local searching and global searching. The food found was the parameter extreme value calculated by Fruit Fly Optimization Algorithm (FOA). Based on the rapid search and fast convergence superiorities of FOA, it is possible to make up the casualness of DPC. An improved fruit fly optimization-based density peak clustering algorithm was proposed as FOA-DPC. The FOA-DPC algorithm would be more efficient and effective than DPC algorithm. The results of seven simulation experiments in UCI data sets validated that the proposed algorithm did not only have better clustering performance, but also were closer to the true clustering numbers. Furthermore, FOA-DPC was applied to practical financial data analysis and the conclusion was also effective.

Keywords: clustering centres; cut-off distance; Density Peak Clustering (DPC); Fruit Fly Optimization

Poboljšana vršna gustoća grupiranja voćnih mušica zasnovana na algoritmu optimizacije i njezina primjena

Izvorni znanstveni članak

Kao algoritam temeljen na gustoći, algoritam grupiranja na osnovu najviše gustoće (Density Peak Clustering - DPC) superioran je u grupiranju pronalazanjem vršne gustoće. No, smanjena udaljenost i središta grupiranja trebaju se postaviti slučajno, što bi utjecalo na rezultate grupiranja. Voćne mušice pronalaze najbolju hranu lokalnim pretraživanjem i globalnim pretraživanjem. Pronađena hrana je ekstremna vrijednost parametra izračunata algoritmom optimizacije voćne mušice (Fruit Fly Optimization Algorithm - FOA). Na osnovu brze pretrage i superiornosti brze konvergencije FOA-e, moguće je nadoknaditi slučajnost DPC-a. Poboljšana vršna gustoća grupiranja voćnih mušica, temeljena na algoritmu optimizacije, predložena je kao FOA-DPC. Taj bi algoritam trebao biti efikasniji i učinkovitiji od DPC algoritma. Rezultati sedam simulacijskih eksperimenata na UCI nizovima podataka potvrdili su da predloženi algoritam nije imao samo bolju performansu grupiranja već je bio bliži pravim brojevima grupiranja. Nadalje, FOA-DPC primijenjen je i u analizi financijskih podataka i pokazao se vrlo učinkovitim.

Ključne riječi: grupiranje vršne gustoće; optimizacija voćnih mušica; skraćena udaljenost; središta grupiranja

1 Introduction

The Clustering technologies become more and more important in digital age. They extract and analyze information and knowledge from increasing data quantity. The clusters identified have high interior similarities and distance difference between clusters [1]. Therefore, the knowledge, pattern and model could easily be found by analyzing the clusters.

Clustering techniques broadly fall into partition clustering, hierarchical clustering, density-based clustering and grid-based clustering [2]. For density-based algorithm, clustering centres were characterized by higher density than their neighbours and relatively large distance from points with higher densities [3]. Density Peaks Clustering (DPC) is one of the typical density-based algorithms, based on the distance between data points [4]. Like DBSCAN [5], the idea of DPC algorithm is to look for the correct number of clusters automatically through detecting nonspherical clusters themselves. DPC adopts the idea of local density maxima from mean-shift [6] and the basic idea of only one parameter of the distance between the data points from K-Medoids [7]. However, there are still some shortcomings in DPC algorithm. The cut-off distance and clustering centres have to be set by personal experience or at random because the cut-off distance influences clustering results. That is to say, the bigger the cut-off distance value, the less clustering number is. It is a challenge to find the optimal results because it totally depends on the personal subjective judgement. For improving DPC algorithm, some scholars had put forward their own ideas [8÷17]. Swarm intelligence algorithm is a good way to optimize the cut-off distance and clustering centres. There are some

meta-heuristic methods, like Particle Swarm Optimization Algorithm (PSO) [18], Artificial Bee Colony Algorithm (ABC) [19], Cuckoo Search Algorithm (CS) [20], Fruit Fly Optimization Algorithm (FOA) [21]. Among these optimization algorithms, FOA and its variants had been proved effective and had been applied to engineering and economics, such as power load forecasting [22], PID controller parameters tuning [23], neural network parameter optimization [24], multi-dimensional knapsack problem [25], steelmaking casting problem [26], and financial distress [27].

Considering the optimization algorithm selected should be more efficient and easy to implementation [28], this paper preferred to select FOA to optimize the cut-off distance and clustering centres in DPC algorithm that had superiority of fast convergence and rapid search. Then FOA-DPC was proposed that have superiority of finding the optimized clustering results quickly.

This paper is organized as follows: Section 2 briefly dwells on DPC algorithm and introduces variables used in the rest of this paper. Section 3 presents FOA idea and general flow. Section 4 introduces the improved algorithm FOA-DPC. Section 5 discusses the simulation experiments and analysis for FOA-DPC. Section 6 extends FOA-DPC to Stock analysis. Section 7 makes some conclusions and remarks.

2 Clustering by fast search and finding of density peaks

In DPC [4], the distance between data points was the basis for this algorithm. It could identify all nonspherical clustering centres and determine the number of clusters. The clustering centres were defined as local maxima and

surrounded by data points in lower density. They were far away from any other clustering centres. There were two important parameters for each data point i . One was the local density ρ_i , and the other was δ_i . The distances d_{ij} among data sets which were assumed to satisfy the triangular inequality. The detail of the DPC was described as follows.

DPC algorithm was only sensitive to ρ_i . The local density ρ_i of data points was defined as:

$$\rho_i = \sum_j \chi(d_{ij} - d_c), \tag{1}$$

δ_i was measured by the minimum distance between the point i and any other point with higher density:

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}). \tag{2}$$

For the data point with highest density, it was taken that $\delta_i = \max_j (d_{ij})$. And d_c had to be chosen to make the average number of neighbours (τ) was around 1 to 2 % of the data set. The point with higher δ_i and higher ρ_i would be chosen by DPC as clustering centre. Then the remaining points were assigned to its nearest neighbour. Detailed calculation methods were the following:

$$\chi(x) = \begin{cases} 1 & x < 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

$$\delta_i = \begin{cases} \max_j (d_{ij}) & \rho = \max(\rho_1, \rho_2, \dots, \rho_n) \\ \min_{j: \rho_j > \rho_i} (d_{ij}) & \text{otherwise} \end{cases}. \tag{4}$$

The core of DPC algorithm was illustrated by the simple example in Fig. 1 and Fig. 2.

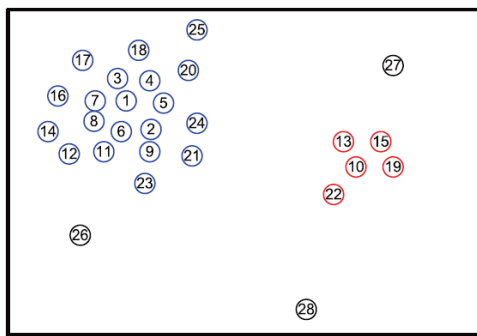


Figure 1 Point distribution in two dimensions

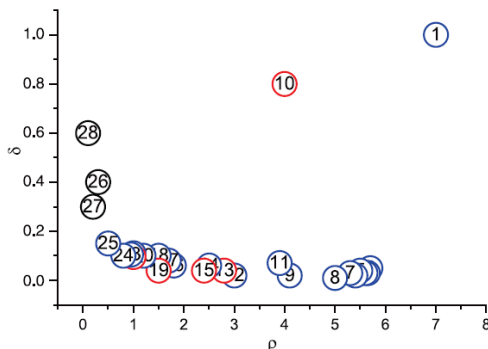


Figure 2 Decision graph in two dimensions

As shown in Fig. 1, there were 28 data points included in two-dimensional space. The density maxima were points 1 and point 10, so the two points were identified as clustering centres. The plots of δ_i shown in Fig. 2 were function of ρ_i for each data point. The points of relatively higher δ and higher ρ were the clustering centres. The points with relatively higher δ and lower ρ were isolated and could be considered as outliers.

3 Fruit Fly Optimization Algorithm

Fruit Fly Optimization algorithm (FOA) was proposed as one of the swarm intelligence algorithms in Science [29], which is based on behaviour of creatures selecting food. The fruit flies had better sensing and perception ability than other birds, so they could find all kinds of scents floating in the air. Guided by the fruit fly with best food flavour, the other flies attached the food selectively, and continued to do the neighbourhood searching process around the attached food. FOA was generated from the activities of a certain amount of fruit fly swarm flying within the search space. Their food finding behaviour had the superiority of fast convergence and rapid search. If FOA reached the maximum times of internal circulation and the quality of food hadn't yet been improved, the corresponding fruit flies started to continue searching randomly. As soon as the number of iterations reached the maximum threshold, the position of higher validity index was the cut-off distance value and the clustering centres would be identified. The algorithm was following:

Step 1: Initialize the fruit fly swarm location

Step 2: Initialize the direction and distance for each fruit fly used to search for favoured food by ophresis.

$$X(i) = X_axis + RandomValue, \tag{5}$$

$$Y(i) = Y_axis + RandomValue. \tag{6}$$

Step 3: If fruit flies could not find the food location, the distance to the origin was set by distance ($Dist_i$), and the next smell concentration judgment value (S_i) was calculated by reciprocal distance.

$$Dist_i = \sqrt{X_i^2 + Y_i^2}, \tag{7}$$

$$S_i = \frac{1}{Dist_i}. \tag{8}$$

Step 4: Substitute smell concentration judgment value (S_i) into smell concentration judgment function to find the smell centre ($Smell_i$) of fruit flies.

$$Smell_i = Function(S_i) \tag{9}$$

Step 5: Find out the fruit fly with maximal smell among the fruit fly swarm.

$$[bestSmell\ bestIndex] = \max(smell) \tag{10}$$

Step 6: Keep the best smell centre value and coordinate x and coordinate y , then the fruit fly swarm fly towards that location by vision.

$$Smellbest = bestSmell \quad (11)$$

$$X_axis = X(bestIndex) \quad (12)$$

$$Y_axis = Y(bestIndex) \quad (13)$$

Step 7: Repeat the implementation of Step 2 to Step 5 and make judgment whether the iterative smell concentration was superior to the previous smell concentration or not. If it was, go back to Step 6.

When the termination criterion was satisfied, then FOA was stopped.

4 Density Peak Clustering Algorithm Based on Fruit Fly Optimization

In DPC algorithm, the cut-off distance and the clustering centres were set by experience or at random. Hence it was hard to find the optimal results because the algorithm needed to set the clustering centres manually. FOA was selected to optimize the cut-off distance and clustering centres in DPC algorithm to make it fast convergent and rapid search. Then FOA-DPC was proposed that has superiority of finding the optimized clustering results quickly. The proposed FOA-DPC algorithm made the ability of self-organization improved, parameters less controlled, and global optimization ability enhanced.

FOA-DPC algorithm had two parameters: the maximum number of iterations T and the population of fruit fly swarm $sizepop$. The flow was as follows:

Step 1: Algorithm initialization. Set d_c value range between (D_{min}, D_{max}) and the *cluster number* range between $(0, C_{max})$. Set the number of fruit flies $sizepop = 10$.

Step 2: Use formula (5), (6) to calculate the value of d_c and *cluster number*;

Step 3: Run DPC algorithm, get n Sil value by (10);

Step 4: Keep the position of food for fruit fly swarm in the search space;

Step 5: If the quality of food had not been improved in continuous limit times, the other fruit flies would start to fly towards the best food and search randomly for next new food;

Step 6: The value of Sil was defined as the objective function. If there is no change in the following time or it reached the maximum iteration number T , it stopped. Output cut-off distance d_c and cluster number according to the best position and clustering results. Otherwise, return to Step 4 to continue updating.

5 Simulation experiments and analysis

Simulation experiment environment is Inter (R) Pentium 2.7 GHz with 4.00 GB memory, 500 G hard disk, Windows 7 system and MATLAB2010a programming language.

5.1 Experimental data

To verify the clustering accuracy of FOA-DPC was higher than that of the DPC algorithm. There were 2 groups of UCI standard data sets and 5 synthetic data sets selected to compare. The experimental data were shown in Tab. 1.

5.2 Evaluating indicator

Silhouette and F-measure indicators were used in this experiment to verify the effectiveness of FOA-DPC algorithm clustering results.

5.2.1 Silhouette indicator

The silhouette indicator is a method of how similar an object is to its own cluster compared to other clusters. A data set D with n sample points was divided into k clusters: $C_i (i = 1, 2, \dots, k)$. $a(t)$ could be the average dissimilarity of sample t in C_j . $D(t, C_i)$ was the average dissimilarity of sample point t to all samples in another cluster C_i , then $b(t) = \min\{d(t, C_i)\}$, wherein $i = 1, 2, \dots, k$ and $i \neq j$. The calculation formula of Sample t Silhouette index Sil was shown in formula (14).

$$Sil(t) = \frac{[b(t) - a(t)]}{\max\{a(t), b(t)\}} \quad (14)$$

$Sil(t)$ value reflected among cluster C_i with compact classes and separable classes. The average of all the samples $Sil(t)$ values reflect the quality of clustering results. The greater the average Sil value, the more compact the class, the better the quality of clustering is.

5.2.2 F-measure indicator

F-measure index was external indicator. It measured a grammar's accuracy, which combines the accuracy rate of $P(i, j)$ with the recall rate of $R(i, j)$. Set up real clustering P_j and clustering C_i . Accuracy rate and recall rate are calculated as Eq. (15) and Eq. (16).

$$P(P_j, C_i) = \frac{|P_j \cap C_i|}{C_i} \quad (15)$$

$$R(P_j, C_i) = \frac{|P_j \cap C_i|}{P_j} \quad (16)$$

F-measure index was calculated by Eq. (17).

$$F(P_j, C_i) = \frac{2 \cdot P(P_j, C_i) \cdot R(P_j, C_i)}{P(P_j, C_i) + R(P_j, C_i)} \quad (17)$$

All F-measure values were weighted average, and the F-measure value of the whole clustering result could be obtained. The larger the F-measure value was, the higher the clustering accuracy was.

$$F = \sum_j \frac{|P_j|}{N} \max_i F(P_j, C_i) \quad (18)$$

5.3 Experimental results and analysis

Comparison of FOA-DPC algorithm with DPC algorithms through simulation experiments showed there were some superiorities of the FOA-DPC algorithm. At first, we used five data sets of UCI that were shown in

Tab. 1 to testify the accuracy (*FM*) of the two clustering algorithms (FOA-DPC and DPC) and the clustering evaluation (*Sil*). The results of the experiment are shown in Tab. 2. Finally, we compared the clustering results of the original DPC, FOA-DPC with the real class number, which are shown in Tab. 3.

Table 1 Experimental data set

Standard data	Sample number	Dimension	Real class number
<i>Iris</i>	150	4	3
<i>Wine</i>	178	13	3
<i>Spiral</i>	312	2	3
<i>Flame</i>	240	2	2
<i>Path based</i>	300	2	2
<i>R15</i>	600	2	15
<i>D31</i>	3100	2	31

Table 2 Comparison of evaluation indexes of clustering results

Standard data	DPC		FOA-DPC	
	<i>Sil</i>	<i>FM</i>	<i>Sil</i>	<i>FM</i>
<i>Iris</i>	0,8462	0,7715	0,7339	0,8407
<i>Wine</i>	0,3277	0,7609	0,4302	0,8121
<i>Spiral</i>	0,3509	0,3820	0,3762	0,3762
<i>Flame</i>	0,4250	1,0000	0,5329	0,9846
<i>Path based</i>	0,5284	0,6415	0,6954	0,6534
<i>R15</i>	0,8985	0,9866	0,9010	0,9932
<i>D31</i>	0,7622	0,9397	0,7633	0,9409

Table 3 Comparison of cluster number

Standard data	DPC	FOA-DPC	Real class number
<i>Iris</i>	3	3	3
<i>Wine</i>	3	3	3
<i>Spiral</i>	3	3	3
<i>Flame</i>	2	2	2
<i>Path based</i>	2	2	2
<i>R15</i>	15	15	15
<i>D31</i>	31	31	31

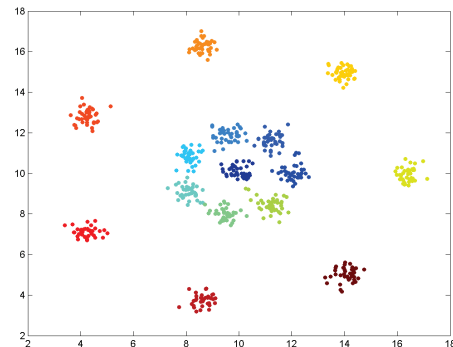
5.4 Comparison of clustering accuracy among different algorithms for different data sets

It could be seen from Tab. 2 that the cluster results of the FOA-DPC algorithms were better than DPC algorithm in the two clustering evaluation indexes. Data sets of *Iris* and *Wine* were selected to test the capability of detecting clustering centres. The *FM* value of FOA-DPC algorithm improved the rate by more than 10 % compared to DPC algorithm. In the *Iris* dataset, the *Sil* value for FOA-DPC algorithm was lower than DPC algorithm. In other datasets, *Sil* values were all increased.

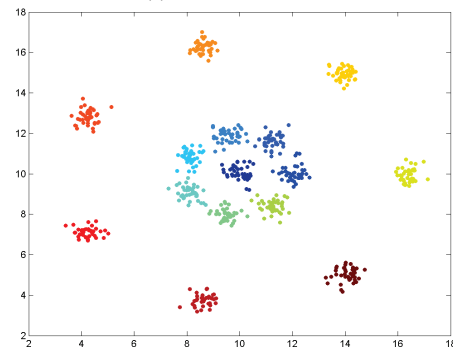
It indicated that FOA algorithm could be combined with DPC algorithm effectively. For data sets of *Spiral*, *Flame*, and *Path based*, the FOA-DPC algorithm also had better performance than DPC algorithm. Especially for *Path based*, the *FM* value of FOA-DPC had significantly improved rate of 30 % compared to DPC algorithm. Meanwhile, in data sets of *D31* and *R15*, it demonstrated that the FOA-DPC algorithm could detect the clustering centres with complicated structure.

Hence, the FOA significantly improved the DPC algorithm which required setting parameters manually. The algorithm based on the behaviours of food searching pattern of fruit flies was used to enhance global optimum ability of DPC algorithm. Silhouette index refers to a

method of interpretation and validation of consistency within clusters of data. This paper introduced the Silhouette index to prevent the algorithm from falling into local extreme and make the algorithm more reasonable to search the optimal solution in the foraging process. So the improved FOA-DPC algorithm enhanced the accuracy of DPC algorithm. This paper used different data sets to compare the clustering outcomes and plot the results of selected UCI data sets in Figs. 3, 4, 5 and 6.



(a) DPC, $d_c = 0,7288$



(b) FOA-DPC, $d_c = 2,2389$

Figure 3 Clustering result of *R15*

5.5 Comparison of clustering results for different data sets

Tab. 3 showed the comparison of cluster results between FOA-DPC and DPC algorithm. As we had seen, FOA-DPC algorithm could find the real classes number of the whole seven data sets. And DPC algorithm obtained the same results. It demonstrated that FOA-DPC could achieve the same clustering results without setting parameters manually, which would reduce the complexity of the algorithm considerably. Especially to the problems with large data, the improved FOA-DPC algorithm would be more efficient and save much more running time. And it confirmed the superiorities of the improved algorithm in Fig. 1, 2, 3, 4, 5. In most test data sets, the FOA-DPC algorithm could be applied to calculate the true cluster numbers. The swarm information was regarded as the knowledge used to guide the searching process. For the three kinds of large data sets *Spiral*, *D31*, and *R15*, FOA-DPC algorithm was just a little improved than the original DPC algorithm.

6 Applications of FOA-DPC on Stock Analysis

In order to test the feasibility of FOA-DPC algorithm, we chose the stock data from SSE (Shanghai Stock Exchange in China). There were four measurement

indexes used in this test: *Net Asset Earning*, *Net Profit Ratio*, *Earning per Share* and *Sales Revenue per Share*. The potential growth ability of listed company was divided into three levels. High level labelled by number 3, which represented Blue Chip Stock. Middle level labelled by number 2 that represented Growth Stocks. Low level labelled by number 1 meant Depressed Stocks. The procedure for applying FOA-DPC algorithm to analysis of the stock data was described as follows:

Step 1: Data pre-processing. The sample stock data were selected from SSE in 2014, which included 44 data points. In this paper, the sample data were normalized to make data in the range from 0 to 1 using the following formula:

$$Y = \{y_i\} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, i = 1, 2, 3 \dots n. \quad (19)$$

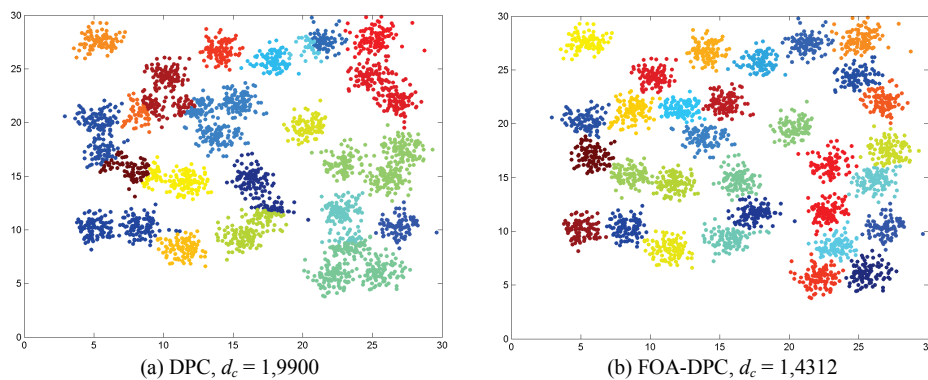


Figure 4 Clustering result of *D3I*

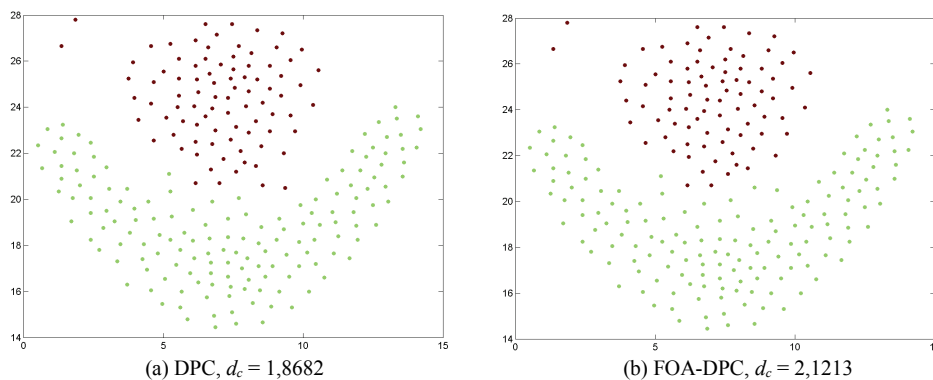


Figure 5 Clustering result of *Flame*

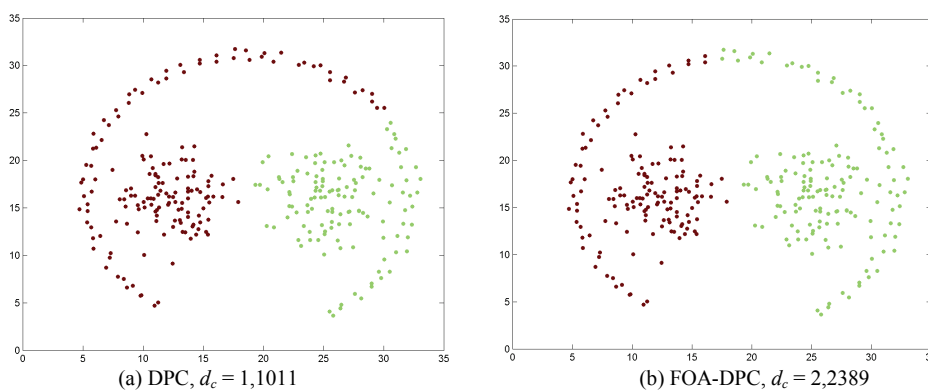


Figure 6 Clustering result of *Path based*

Step 2: Train FOA-DPC Algorithm. In the FOA-DPC Algorithm, the cut-off distance value of the DPC Algorithm was dynamically tuned by the FOA to calculate values of d_c and *cluster number*. By simulation, the d_c value and *cluster number* value were 0,5809 and 3 respectively. The final clustering results are shown in Tab. 4.

As seen in Tab. 4, the actual level and test level were basically identical. The net asset earning and earning per share were both the most important indexes to measure

earning power. Level 1 represented the growth stocks, which had ability to make profits except growing latent. Level 2 meant the stationary stocks and developed smoothly. Level 3 showed the blue chip stocks. The net asset earnings and earnings per share in level 3 were apparently higher than the other two categories. The stock analysis showed that the application of the proposed algorithm on stock analysis was effective. Hence, it provided a relatively effective research tool for the stack classification field.

Table 4 Results of the FOA-DPC clustering

Stock code	Net asset earning (CNY)	Net profit ratio (%)	Earning per share (CNY)	Sales revenue per share (CNY)	Level (actual)	Level (test)
000952	-43,900	-45,255	-0,869	1,930	1	1
600129	-27,300	-3,970	-0,648	16,299	1	1
000606	-14,400	-32,210	-0,245	0,760	1	1
600789	-8,880	-5,410	-0,215	3,976	1	1
000788	-2,440	-1,190	-0,045	3,802	1	1
600671	3,440	1,820	0,022	1,221	2	2
600466	2,100	3,150	0,029	0,909	2	2
600222	3,000	2,550	0,065	2,526	2	2
600829	1,900	2,290	0,069	2,998	2	2
002099	1,680	4,050	0,074	1,824	2	2
300110	3,790	6,120	0,083	1,363	2	2
002118	2,360	6,080	0,092	1,512	2	2
600645	3,430	7,430	0,101	1,353	2	2
600080	3,030	4,340	0,101	2,335	2	2
000756	2,780	1,410	0,111	7,850	2	2
600664	2,990	1,490	0,129	8,610	2	2
000153	3,680	2,410	0,131	5,443	2	2
300158	2,140	2,180	0,146	6,665	2	2
002365	2,860	4,810	0,170	3,521	2	2
000790	10,450	12,510	0,172	1,373	2	3
300194	2,670	10,180	0,176	1,724	2	2
600216	2,590	3,510	0,181	5,162	2	2
000597	3,260	2,000	0,183	9,129	2	2
300086	3,380	14,460	0,197	1,362	2	2
002349	2,850	6,830	0,197	2,889	2	2
600351	9,380	9,000	0,246	2,731	2	3
600513	10,130	6,570	0,281	4,267	2	3
000566	10,010	12,110	0,329	2,713	2	3
300147	11,280	12,910	0,386	2,991	2	3
002412	10,500	15,940	0,397	2,491	2	3
002393	3,190	12,000	0,509	4,241	3	2
600993	12,820	12,430	0,608	4,888	3	3
002317	12,140	17,160	0,608	3,541	3	3
600557	12,780	12,470	0,622	4,989	3	3
300049	10,440	13,510	0,672	4,972	3	3
300255	9,760	16,480	0,710	4,306	3	3
603998	10,020	19,480	0,743	3,812	3	3
600976	11,930	6,610	0,756	11,431	3	3
002550	11,670	30,080	0,766	2,546	3	3
002728	10,900	23,260	0,802	3,447	3	3
300233	9,730	10,340	0,861	8,320	3	3
600750	12,740	9,340	0,883	9,447	3	3
300294	12,130	23,790	1,374	5,776	3	3

7 Conclusion

On the basis of FOA and DPC algorithm, an improved clustering algorithm was proposed as FOA-DPC. For DPC algorithm, the cut-off distance and clustering centres had to be set by experience or at random. Furthermore, the clustering results of DPC algorithm are sensitive to different parameters. In order to alleviate the parameters-sensibility of density peak clustering, FOA-DPC algorithm computed the cut-off distance value and clustering centres by introducing the global searching ability into DPC. We had chosen *Sil* index as the objective function, hence two parameters found by FOA were the extreme value in solution space.

The simulation results showed that FOA-DPC algorithm was more accurate to calculate the true number of clusters than DPC algorithm. Also it demonstrated that the number of clusters was consistent with the real cluster numbers of FOA-DPC without setting parameters

manually. The experiments exhibited the advantages of FOA-DPC algorithm. This superiority reduced the complexity of the algorithm considerably, especially in solving the problems with large scale data. The improved FOA-DPC algorithm proved to be more efficient and could save computing time. The clustering performance and the evaluation results were both significantly improved. FOA-DPC algorithm was then applied to real data analysis. Through analyzing the data from Shanghai Stock Exchange in China, the algorithm effectiveness was proved effective again. However there were still some difficulties in distinguish boundary nodes and low density nodes, which are our future work.

Acknowledgements

The research was supported by the National Science Foundation of China under grant No. 61202306, 61472049, 61402193, 61572225 and 61202309; the

National Society Science Foundation of China under grant No. 15BGL090; the Foundation of Jilin Provincial Science & Technology; Department under Grant No. 201215119, 20130522177JH, 20130101072JC, 20150312001ZG and 20160101259JC; the Society Science Research Foundation of Jilin Province under grant No. 2014B166 and 2015BS46; Jilin Province Education Department under grant No. 2014159 and 2016 [78]; the Open Fund Project of Logistics Industry Economy and the Intelligent Logistics Laboratory of Jilin Province under grant No. 02060311 and 02060312; the Fund of Jilin Province Economic Structure Strategic Adjustment under grant No. 2015Y040 and 2015Y037.

8 Reference

- [1] Du, M.; Ding, S.; Jia, H. Study on density peaks clustering based on k-nearest neighbors and principal component analysis. // *Knowledge-Based Systems*. 135, 2(2016), pp. 135-145. DOI: 10.1016/j.knsys.2016.02.001
- [2] Han, J. W.; Damber, M. *Data Mining Concepts and Techniques*. 2nd ed., New York: Elsevier Inc, 2006, pp. 384-424.
- [3] Frey, B. J.; Dueck, D. Clustering by Passing Messages Between Data Points. // *Science*. 315, 5814(2007), pp. 972-976. DOI: 10.1126/science.1136800
- [4] Alex, R.; Alessandro, L. Clustering by fast search and find of density peaks. // *Science*. 344, 6(2014), pp. 1492-1496.
- [5] Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. // *Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.
- [6] Cheng, Y. Mean shift, mode seeking, and clustering. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 12, (1995), pp. 790-799. DOI: 10.1109/34.400568
- [7] Kaufman, L.; Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*. // *Journal of the American Statistical Association*. 1990. DOI: 10.1002/9780470316801
- [8] Du, M.; Ding, S.; Jia, H. Study on density peaks clustering based on k-nearest neighbors and principal component analysis. // *Knowledge Based Systems*. 99, (2016), pp. 135-145. DOI: 10.1016/j.knsys.2016.02.001
- [9] Wang, M.; Zuo, W.; Wang, Y. An improved density peaks-based clustering method for social circle discovery in social networks. // *Neurocomputing*. 179, (2016), pp. 219-227. DOI: 10.1016/j.neucom.2015.11.091
- [10] Zhang, W.; Li, J. Extended fast search clustering algorithm: widely density clusters, no density peaks. // *Computer Science*. (2015).
- [11] Aksehirli, E.; Goethals, B.; Muller, E. Efficient cluster detection by ordered neighborhoods. // *In Big Data Analytics and Knowledge Discovery*, 2015, pp. 15-27.
- [12] Chen, Y.; Lai, D.; Qi, H.; Wang, J.; Du, J. A new method to estimate ages of facial image for large database. // *Multimedia Tools and Applications*. (2015), pp. 1-19.
- [13] Li, Y.; Liu, W.; Wang, Y.; Tao, D. Co-spectral clustering based density peak. // *In 2015 IEEE 16th International Conference on Communication Technology (ICCT)*, 2015, pp. 925-929.
- [14] Wiwie, C.; Baumbach, J.; Rottger, R. Comparing the performance of biomedical clustering methods. // *Nature Methods*. 12(2015), pp. 1033-1038. DOI: 10.1038/nmeth.3583
- [15] Sun, K.; Geng, X.; Ji, L. Exemplar component analysis: A fast band selection method for hyperspectral imagery. // *IEEE Geoscience and Remote Sensing Letters*. 12(2015), pp. 998-1002. DOI: 10.1109/LGRS.2014.2372071
- [16] Jia, S.; Tang, G.; Hu, J. Band selection of hyperspectral imagery using a weighted fast density peak-based clustering approach. // *In Intelligence Science and Big Data Engineering Image and Video Data Engineering*, 2015, pp. 50-59.
- [17] Xie, K.; Wu, J.; Yang, W.; Sun, C. K-means clustering based on density for scene image classification. // *In Proceedings of the 2015 Chinese Intelligent Automation Conference*, 2015, pp. 379-386. DOI: 10.1007/978-3-662-46469-4_40
- [18] Kennedy, J.; Eberhart, R. *Particle swarm optimization*. // *IEEE Press*. (1995), pp. 942-1948. DOI: 10.1109/icnn.1995.488968
- [19] Karaboga, D. An idea based on honey bee swarm for numerical optimization. // *Technical report*, Erciyes University, 2005.
- [20] Yang, X.; Deb, S. Cuckoo search via Levy flights, *Proceedings of the World Congress on Nature and Biologically Inspired Computing*. // *IEEE*. 12(2009), pp. 210-214.
- [21] Pan, W. A new evolutionary computation approach: fruit fly optimization algorithm. // *Conference of Digital Technology and Innovation Management*, 2011, Taipei.
- [22] Li, H.; Guo, S.; Li, C.; Sun, J. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. // *Knowl. Based Systems*. 37, (2013), pp. 378-387. DOI: 10.1016/j.knsys.2012.08.015
- [23] Han, J.; Wang, P.; Yang, X. Tuning of PID controller based on fruit fly optimization algorithm. // *International Conference on Mechatronics and Automation (ICMA)*. (2012), pp. 409-413. DOI: 10.1109/icma.2012.6282878
- [24] Pan, W. A new fruit fly optimization algorithm: taking the financial distress model. // *Knowl. Based Systems*. 26, (2012), pp. 69-74. DOI: 10.1016/j.knsys.2011.07.001
- [25] Wang, L.; Zheng, X.; Wang, S. A novel binary fruit fly optimization algorithm for solving the multidimensional knapsack problem. // *Knowl. Based Systems*. 48, (2013), pp. 17-23. DOI: 10.1016/j.knsys.2013.04.003
- [26] Li, J.; Pan, Q.; Mao, K.; Suganthan, P. Solving the steelmaking casting problem using an effective fruit fly optimisation algorithm. // *Knowl. Based Systems*. 72, (2014), pp. 28-36. DOI: 10.1016/j.knsys.2014.08.022
- [27] Pan, W. Mixed modified fruit fly optimization algorithm with general regression neuralnetwork to build oil and gold prices forecasting model. // *Kybernetes, Emerald*. 43, 9(2014), pp. 1053-1063. DOI: 10.1108/K-02-2014-0024
- [28] Pan, W. *Fruit Fly Optimization Algorithm*. // Tsang Hai Book Publishing Co., Taiwan-Taichung, 2011.
- [29] Pan, W. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. // *Knowledge-Based Systems*. 26, 7(2011), pp. 69-74.

Authors' addresses

Ruihong Zhou, postgraduate, lecturer
School of Management, Jilin University,
School of Management Science and Information Engineering,
Jilin University of Finance and Economics,
3699 Jingyue Street, Changchun 130117, China
zrh@jlufe.edu.cn

Qiaoming Liu, postgraduate
School of Computer Science and Engineering,
Changchun University of Technology,
2055 Yanan Street, Changchun 130012, China
lqmmring@163.com

Zhengliang Xu, professor*Corresponding author*

School of Management, Jilin University,
2699 Qianjin Street, Changchun 130012, China
zlxu@jlu.edu.cn

Limin Wang, professor

School of Management Science and Information Engineering,
Jilin University of Finance and Economics,
3699 Jingyue Street, Changchun 130117, China
wlm_new@163.com

Xuming Han, professor

School of Computer Science and Engineering,
Changchun University of Technology,
2055 Yanan Street, Changchun 130012, China
hanxuming@163.com