IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-DRIVEN MONITORING, FAULT DIAGNOSIS AND CONTROL OF CYBER-PHYSICAL SYSTEMS

Received November 24, 2017, accepted January 1, 2018, date of publication January 23, 2018, date of current version April 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2791203

Multi-Parameter-Setting Based on Data Original Distribution for DENCLUE Optimization

YI LUO[1], KE ZHANG[2], YI CHAI[2], AND YINGZHI XIONG[3]
[1]School of Automation, Chongqing University, Chongqing 400030, China
[2]State Key Laboratory of Power Transmission Equipment and System Security and New Technology, Chongqing University, Chongqing 400030, China
[3]China Automotive Engineering Research Institute Co., Ltd., Chongqing 401122, China

Corresponding author: Ke Zhang (smeta@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61673076, Grant 61203084, and Grant 61633005.

ABSTRACT DENCLUE is a typical density-based clustering method, and it also is an important pattern classification analysis technique. In this clustering method, the propriety of parameters' values greatly influences the quality of distinguishing. Accordingly, how to choose an appropriate value of every parameter is a problem that is worth studying. The problem is focused in this paper, and a method which is very different from previous ones is presented. The highlight of the method is that the selection of parameters no longer depends on personal experience but on data original distribution. More specifically, smoothing parameter h is more or less proportional to the average value of distance between two arbitrary data points; step size δ is adapted according to the density of data points in the hill-climbing progress; noise threshold ξ is replaced by δ and total number of data points. Compared with the original DENCLUE and an improved DENCLUE proposed in our previous research, the optimized algorithm can bring better clustering from experiments. In addition, due to the adaptiveness of δ, the method will become less complex.

INDEX TERMS DENCLUE, multi-parameter-setting, optimized, data distribution, step size adaptive.

I. INTRODUCTION

Clustering is a main task of data mining. It groups a set of objects in such a way that in the same group (called a cluster), objects are more similar to each other than to those in other groups [1]. This technique is widely used in machine learning, pattern recognition, image analysis, data compression, and so on.

Generally, ordinary clustering methods can be distributed into 5 classes: partitioning-based clustering, hierarchical-based clustering, grid-based clustering, model-based clustering and density-based clustering [2]–[4]. The difference in the first 4 methods is that in density-based clustering, the distance between two data points is substituted by distribution density of data points as principle for clustering. The advantage of density-based clustering is that clustering results can reflect the distribution of data points objectively. In practice, the shape of a cluster is very flexible, which means different clusters have different responding shapes. In density-based clustering, data points in dense area can be found and classified into corresponding clusters, so the formations of clusters are not influenced exclusively by the distance between data points [5].

DENCLUE is an efficient density-based clustering. In this method, a data point i is influenced by other data points. The influence can be described by a density estimation function which is a sum of kernel functions of corresponding data points for i. With the help of DENCLUE, some special data sets (for example, unevenly-distributed data points or shape-irregular clusters etc.) can be well grouped. As we known, there are 3 parameters h, δ, ξ called smoothing parameter, step size and noise threshold, respectively. The 3 parameters can influence the quality of clustering. Only when the selection of parameters' values is appropriate can the clustering result be consistent with original data. Unfortunately, in this method, the 3 parameters' setting depends on personal experience. If rich and powerful prior knowledge is lacked, it is difficult to select appropriate values of the 3 parameters. In addition, some processes in the method cost too much time.

For these problems, some improved methods are proposed. Hinneburg and Gabriel [6] proposed a method to adjust the value of step size δ so that the extreme point will not be missed in hill-climbing process. In order to magnify the similarity between different points, Suganya and Nagarajan [7] proposed a DENCLUE method based on fuzzy theory.

Chang *et al.* [8] and Guo and Zang [9] presented methods without noise threshold by introducing the niching genetic algorithm and nichePSO respectively. In [10], for the problem of how to evaluate the similarity between different information in network, K.Santhi Sree raised a method based on page reference sequence and the page information. For the real-time fault diagnosis, Zhu *et al.* [11], introducing substituting distance coefficients with similarity coefficients, solved the problem that the DENCLUE is not stabilized for different data sets. Yan *et al.* [12] defined a new density estimation function which can better describe the effect of all the data points on a specific data point. Zhang *et al.* [13] proposed a method in which noise threshold is replaced by step size and data point number in the set. The algorithm based on avoid determining noise threshold in DENCLUE will be refferred to simply as ABADNT in this paper. In order to speed the calculation of DENCLUE, Rehioui *et al.* [14] avoids the crucial step in hill-climbing process. In [15], hill-climbing process was repalced by Simulated Annealing (SA) and by Genetic Algorithm. He and Pan [16] used dynamic noise threshold to replace fixed noise threshold in original DENCLUE.

In our previous research, the clustering analysis method has been applied in fault diagnosis many times. We propose a class of category discrimination based model frameworks for multiple fault diagnosis [17]–[19]. In these frameworks, density-based clustering plays an important role in fault diagnosis, feature extraction and fault isolation. But in practice, fault diagnosis is influenced by initialied values of several parameters, so the result is not stable. Therefore, we hope that we can get a method for clustering parameter setting in which human intervention is avoided, so that this method can satisfy the needs of various engineering applications in the absence of sufficient prior knowledge.

The study in this paper is based on abundant previous related researches. In our previous research, we presented ABADNT which avoids Determining Noise Threshold. As a subsequent research, smoothing parameter h and step size δ are determined by the original distribution of data in data set, and the function of noise threshold ξ is replaced by that of data point number in the set and step size δ. This method can avoid manually adjusting parameters.

The paper is organized as follows. In Section 2, existing problems in DENCLUE will be analyzed. In Section 3, we will propose an improved algorithm to avoid the use of artificially-set parameters. In Section 4, we will compare the clustering results of our method with those of the original DENCLUE and ABADNT. Section 5 will contain the conclusion and future work.

II. EXISTING PROBLEMS IN DENCLUE

In DENCLUE, the effect of a data point j on another data point i can be described by a kernel function. In addition, the total effect of all data points on i can be described by a density estimation function which is a sum of kernel functions of corresponding data points for i. In this way, the influence of each point data on the whole data set can be described intuitively, and the data point distribution which is unknown in advance in a set can be reflected objectively [20], [21] by density estimation function.

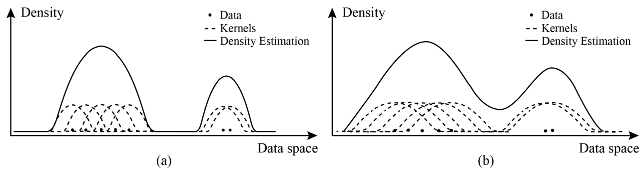

FIGURE 1. Kernel density estimation curve with different smoothing parameters: (a) smaller *h*; (b) bigger *h*.

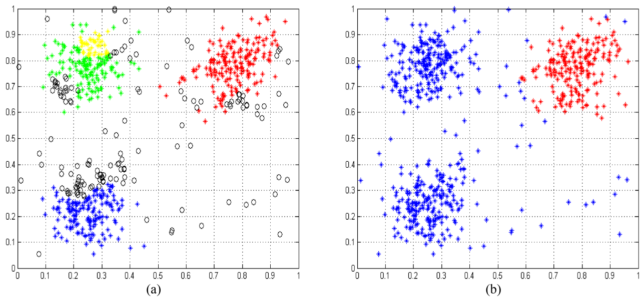

FIGURE 2. The influence of different noise threshold on clustering results: (a) noise threshold is too small; (b) noise threshold is too large.

In the original DENCLUE, there are 3 parameters that should be set. They are smoothing parameter h, noise threshold ξ and step size δ.

Steep degree of kernel function is affected by smoothing parameter h. The smaller h is, the more steep the kernel function is, and vice versa [6] (see Fig.1). In other words, if h is small, when the distance between data i and data j becomes lager, the effect of j on i will decrease rapidly; if h is big, when the distance between data i and data j becomes larger, the effect of j on i will decrease slowly. What's more, the steep degree of density estimation function is also affected by smoothing parameter h. The smaller h is, the more steep the density estimation function is, and vice versa [6] (see Fig.1).

Noise threshold ξ will decide whether a data point is a noise point or not, and ξ is also a decisive factor of whether two clusters can be combined to one cluster. If noise threshold is too big, some points in the cluster may be considered as noise points, and the same class may be wrongly divided into two clusters. On the contrary, if noise threshold is too small, noise points may be mistaken as the points inside the cluster, and clusters that are not of the same cluster may be classified into one category. Fig. 2 shows that the distribution of data points in set is a Gaussian distribution. In Fig. 2 (a), we can see that when noise threshold ξ is too small, many data points which belong to clusters are wrongly considered as noise points, and one cluster is divided into two clusters. In Fig. 2 (b), we can see that when noise threshold ξ is too large, many noise points are considered as data points belonging to clusters, and two different clusters are wrongly combined into one cluster.

Step size δ affects step length for finding density attractor. If the δ is too big which results in a too large step, we may miss the density attractor when we try to find a local extreme point through hill-climbing process. If the δ is too small which results in a too small step, there may be too many steps before finding the density attractor.

From the analysis above, we can see that the 3 parameters are very important for the DENCLUE. In the original DENCLUE, appropriate selection of parameter values depends on personal experience. In some applications, because of the lack of rich and powerful prior knowledge, artificial parameter deviation will make clustering process longer and even lead to that the results can not be achieved. In order to solve this problem, we will propose an improved DENCLUE. The core insight is to avoid frequent artificial adjustment of smoothing parameter h, noise threshold ξ and step size δ. Values of the parameters are determined by the distribution of data sets.

III. MULTI-PARAMETER-SETTING BASED ON DATA ORIGINAL DISTRIBUTION FOR DENCLUE OPTIMIZATION

In the beginning of this section, we make several definitions as follows.

In order to describe the effect of data point j on data point i, the influence function with a Gaussian kernel is defined as follows:

$$f_j(i) = e^{-\frac{d(i,j)^2}{h^2}} \tag{1}$$

Suppose there are N data points in the data set. For any data point i, d^i_{Min} stands for the distance between data point i and the closest data point from i.

$$d^C_{Ave} = \frac{1}{N_C}\sum_{i=1}^{N_C} d^i_{Min} \tag{2}$$

where d^C_{Ave} stands for the average value of d^i_{Min} of all the data points in the cluster C.

$$d_{Ave} = \frac{1}{N}\sum_{i=1}^{N} d^i_{Min} \tag{3}$$

where d_{Ave} stands for the average value of d^i_{Min} of all the data points in the data set.

$$\sigma(d)^2 = \frac{1}{N}\sum_{i=1}^{N} (d_i - d_{Ave})^2 \tag{4}$$

where $\sigma(d)^2$ stands for the variance of d^i_{Min} of all the data points in the data set.

A. DEFINITION FOR SMOOTHING PARAMETER h

The steep degree of density estimation function is affected by smoothing parameter h. For different data sets, the most appropriate value of h is not the same. Suppose $h = C$. If d^i_{Min} of any data point in the data set meets the condition: $d^i_{Min} >> C$, the effect of any other data points on i is very small; if d^i_{Min} of any data point in the data set meets the condition: $d^i_{Min} << C$, the effect of any other data points on i can not be ignored. In the former situation, the number of the clusters may be far more than the original source; in the later situation, the number of the clusters may be less than the original source.

In this paper, we consider that smoothing parameter h is more or less proportional to d_{Ave}. Fig. 3 shows the values of density estimation of all the coordinate points in the data space when h equals to 0.5,$\sqrt{2}$, 2 and $\sqrt{10}$, respectively.

From the distribution of the data set, the data points can be classified into 4 clusters intuitively: the cluster in the left bottom corner including 14 points, the cluster in the left top corner including 8 points, the cluster in the middle including 5 points and the cluster in the right top corner including 3 points. The isolated data point in the middle top can be seen as a noise point. The d^i_{Min} of the point located in (1, 7) is the smallest, which equals to 0.25. The d^i_{Min} of the point located in (7, 8) is the biggest, which equals to $\sqrt{10}$.

Fig. 3 (a) shows that there are multiple extreme points, so the cluster in the right top corner and the cluster in the middle may be wrongly divided into multiple clusters in the subsequent process.

From Fig. 3 (d), we can see that there is only one extreme point, so all the data points may be classified into one cluster.

From both Fig. 3 (b) and Fig. 3 (c), we can see that there is, and there only is, one extreme point around each cluster, so the data points can be classified into corresponding clusters.

But there is a little difference between Fig. 3 (b) and Fig. 3 (c). In the Fig. 3 (b), there is an extreme point around the noise point which can be classified alone into a cluster and be judged as a noise point in the subsequent process. In the Fig. 3 (c), there is not an extreme point around the noise point. The density attractor of the noise point will move along the fastest-changing direction (the red arrow) of the density estimation to the area around the coordinate point (6.5, 5.2). Finally, the noise point may be wrongly classified into the cluster in the middle. The reason for this result is that the distance between noise point and the data point (6, 5) is $\sqrt{10}$ which is not much bigger than h($h = 2$ in Fig. 3 (c)). So, the effects of data points in the cluster on noise point make the density attractor of the noise point move to the cluster.

From the above analysis, a definition for the value of h is proposed. Suppose there are M clusters in a data set. They are $C_1, C_2, C_3, \ldots, C_M$ and $d = \min\left\{d^{C_1}_{Ave}, d^{C_2}_{Ave}, \ldots, d^{C_M}_{Ave}\right\}$, $D = \max\left\{d^{C_1}_{Ave}, d^{C_2}_{Ave}, \ldots, d^{C_M}_{Ave}\right\}$, the h is

$$h = \sqrt{d^2 + D^2} \tag{5}$$

B. AVOID DETERMINING NOISE THRESHOLD ξ

In the original DENCLUE, noise threshold ξ mainly functions in two aspects:

(1) Distinguish the noise points from all the data points in data set.

(2) Judge whether two clusters where there exists one density attractor respectively can be combined into one cluster.

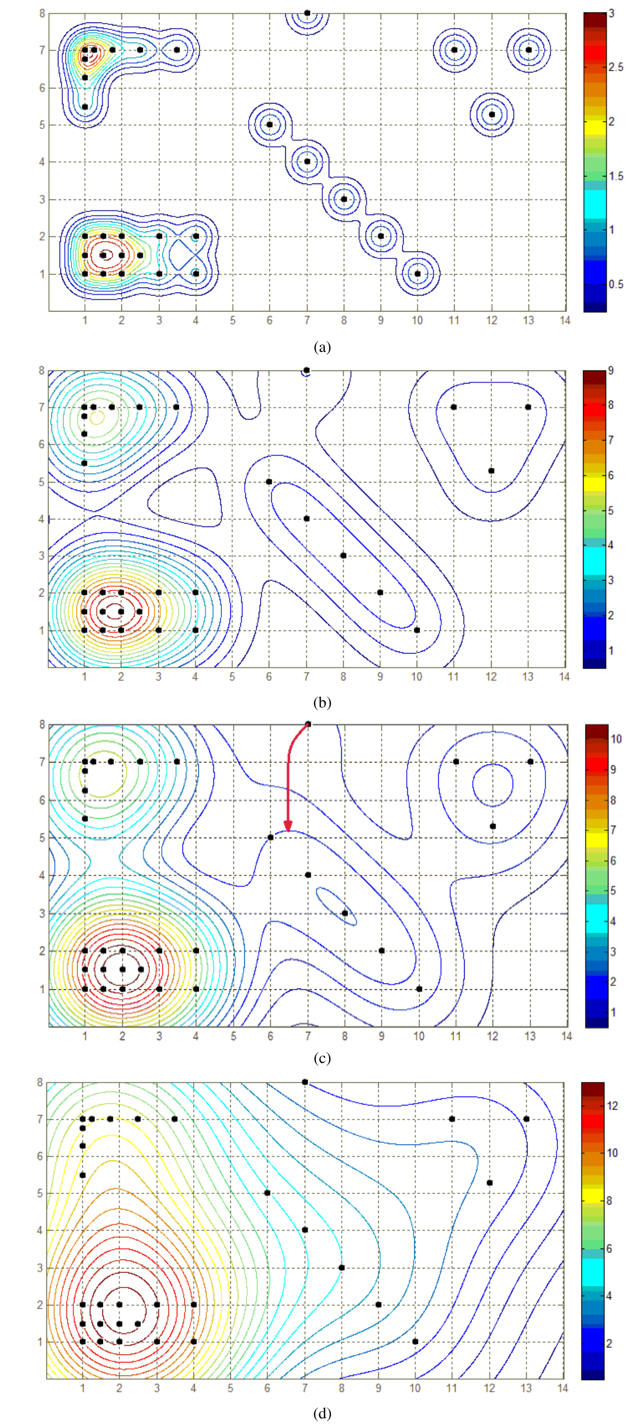

FIGURE 3. **The values of density estimation of all the coordinate point in the data space. (a) $h = 0.5$; (b) $h = \sqrt{2}$; (c) $h = 2$; (d) $h = \sqrt{10}$.**

In order to avoid determining noise threshold ξ, we must have some methods to solve two problems above without ξ.

Noise is, by definition, random error or variance of measured variables, and has significant difference with data points within the cluster [22]. Normally, in each data set, the number of noise points is much less than the regular data points. If the noise points are divided into clusters wrongly, their number should be no greater than that of regular points in clusters containing the minimum number of data points. So, noise identification can be converted into point number determination within the cluster [23]. When noise points are distributed evenly and the number of noise points is far smaller than the number of points in clusters, experience-based method can be applied. For data set with n points, the number of cluster is about $\sqrt{0.5n}$. Each cluster has $\sqrt{2n}$ points. So, in the practice, if there are more than $\sqrt{2n}$ points in a set, this set should be a cluster. Otherwise, the points in the set are noise points.

The other function of noise threshold ξ is judging whether two clusters where there exists one density attractor respectively can be combined into one cluster. To address this problem, the author of DENCLUE proposed a method to judge whether two clusters can be combined to one cluster by analyzing step size δ [6]. Set a small δ as cluster distance threshold. Two clusters where there exists one density attractor respectively can be combined into one cluster if the distance between the two density attractors is less than 2δ.

From the analysis above, noise threshold ξ can be replaced by inner-cluster point estimation and step size δ.

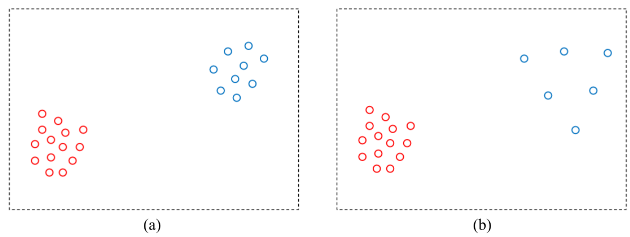

FIGURE 4. **The distributions of two data sets: (a) data set with small $\sigma(d)^2$; (b) data set with big $\sigma(d)^2$.**

C. AN ADAPTIVE METHOD FOR STEP SIZE δ

In the original DENCLUE, in order to find the density attractor of data point x, we will use hill-climbing process shown as follows:

$$x^0 = x \tag{6}$$

$$x^{i+1} = x^i + \delta \frac{\nabla \hat{f}\left(x^i\right)}{\left\| \nabla \hat{f}\left(x^i\right) \right\|} \tag{7}$$

where $\nabla \hat{f}\left(x^i\right)$ is the gradient of $\hat{f}\left(x^i\right)$, we have

$$\nabla \hat{f}\left(x^i\right) = \sum_{k=1}^{N} \left(x_k - x^i\right) \cdot f^{x_k}\left(x^i\right) \tag{8}$$

In the hill-climbing process, if $\hat{f}\left(x^{i+1}\right) \leq \hat{f}\left(x^i\right)$, x^i is a local maximum point, namely a density attractor of x. Thus, stop climbing process.

From the hill-climbing process in the original DENCLUE, we can see that for any data point, step size δ is a constant value. In Fig. 4 (a), there are two clusters: blue and red. The density of data points in blue is basically equivalent to that

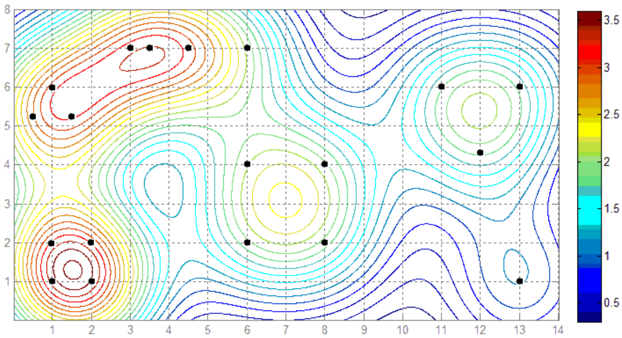

FIGURE 5. The values of density estimation of all the coordinate point in the data space.

of data points in red. In other words, the $\sigma(d)^2$ for this data set is not big. In this situation, δ being a constant value is not a problem for hill-climbing process. In Fig. 4 (b), the two clusters are also blue and red. The density of data points in blue is lower than that of data points in red. In other words, the $\sigma(d)^2$ for this data set is big. In this situation, if the value of δ is appropriate for finding the density attractor of points in blue, it may be too big so that the density attractor of points in red may be missed; if the value of δ is appropriate for finding the density attractor of points in red, it may be too small so that finding the density attractor of points in blue will cost too much time.

In order to solve this problem, an adaptive method for step size δ is proposed in the paper. In the Fig. 5, the black dots stand for data points, and the contour lines represent the density estimation of coordinate points. We can see that the extreme points appear around a data point or around the geometrical center of some data points. In addition, we can find a relation: $D < d^i_{Min}$(D is the distance between an extreme point and i, the closest data point from the extreme point.).

We propose an assumption: in the progress of finding the density attractor of point x, if the distance D between x^k and i (i is the closest data point from x^k) is less than d^i_{Min}, that step size δ equals to $0.5d^i_{Min}$ can not only guarantee the density attractor not to be missed, but also accelerate the progress.

Based on the assumption, the hill-climbing process with adaptive step size δ is shown as follows: x is any data point in the data set.

$$x^0 = x \tag{9}$$

$$\delta^0_x = 0.5 \times d^x_{Min} \tag{10}$$

$$x^{k+1} = x^k + \delta^k_x \frac{\nabla \hat{f}\left(x^k\right)}{\left\|\nabla \hat{f}\left(x^k\right)\right\|} \tag{11}$$

Find the data point i which is the closet point from x^{k+1}, and upgrade step size δ^{k+1}_x. $\mathrm{h} = (\mathrm{d}^2 + \mathrm{D}^2)^{0.5}$

$$\delta^{k+1}_x = 0.5 \times d^i_{Min}. \tag{12}$$

In the hill-climbing process, if $\hat{f}\left(x^{k+1}\right) \le \hat{f}\left(x^k\right)$, x^k is a local maximum point, namely density attractor of x. Thus, stop climbing process.

Now, the last problem is how to judge whether two clusters where there exists one density attractor respectively can be combined into one cluster or not. Using Alexander Hinneburg's thought, we propose the following method.

Suppose there are N data points in the data set. When we get the density attractor of point i by hill-climbing process with adaptive step size δ, the last step size for i is δ^{final}_i. For all the data points, we can get a last-step-size set $\left\{\delta^{final}_1, \delta^{final}_2, \cdots, \delta^{final}_N\right\}$. Firstly, data points having a same density attractor are classified into a same cluster. Then calculate the distances between all of the density attractors. For two arbitrary density attractors i and j, if the distance between them is less than $2\delta^{final}_i$ or $2\delta^{final}_j$, the 2 clusters to which i and j belong respectively can be combined into one cluster. This process will not stop until there are no clusters to be combined.

D. AN OPTIMIZED DENCLUE

With the help of above work in this section, multi-parameter-setting based on data original distribution for DENCLUE optimization can be proposed (the sketch map shown in Fig.6). The specific steps are shown as follows:

(1) Calculate the d^i_{Min} for each data point in data set, and sort the results from smallest to biggest. d stands for the average value of top smallest 10% values of d^i_{Min}. D stands for the average value of top smallest 85%-95% values of d^i_{Min} (the corresponding data points of top biggest 5% values of d^i_{Min} are seen as noise points in this paper).

(2) For each data point x, find its density attractor by the hill-climbing process with adaptive step size δ. (using the equations (9), (10), (11) and (12).)

(3) Firstly, data points having a same density attractor are classified into a same cluster. Then calculate the distances between all of the density attractors. For two arbitrary density attractors i and j, if the distance between them is less than $2\delta^{final}_i$ or $2\delta^{final}_j$, the two clusters to which i and j belong respectively can be combined into one cluster. This process will not stop until there are no clusters to be combined.

(4) Judge whether the number of points in each set is greater than $\sqrt{2n}$ (n is the total number of data points). If it is, the set is considered as a cluster; if not, all points in the set are considered as noise points.

IV. SIMULATION AND DISCUSSION

In this section, our method will be compared with the original DENCLUE. The separation between different clusters and the compactness within a cluster are two factors to evaluate the qualities of clustering results. So, in this paper, silhouette coefficient will be an evaluation function [24], [25]. Suppose there are n data points in data set D which can be divided into k clusters: $C_1, C_2, \ldots, C_k$. For any data point i, $a(i)$ represents average distance between i and other data points

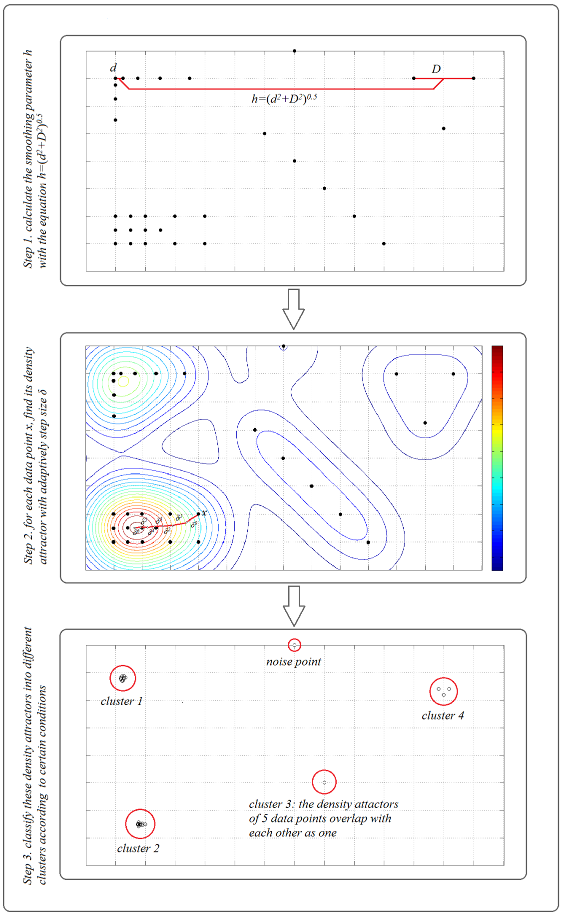

FIGURE 6. **The sketch map of optimized DENCLUE.**

in the same cluster. $b(i)$ represents the minimum average distance between i and data points which does not belong to i's cluster but are in the same cluster. The silhouette coefficients is defined as follows:

$$s(x) = \sum_{i=1}^{n} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{13}$$

where $s(x)$ is the average value of silhouette coefficients for all points. The smaller $a(i)$ is, the more compact the cluster is, and vice versa. The bigger $b(i)$ is, the more detached i and data points in any other clusters are, and vice versa. So, if $s(x)$ is big, the data points in the same clusters are compact, but the data points in different clusters are detached.

In order to evaluate the quality of clustering result of our method, 3 data sets [12] are used. They are called Three Circles, Two Moons, and Linear Segmentation. Three Circles is composed of 3 concentric circles. This data set is characterized by flow pattern, embedded structure, and density difference. Two Moons is composed of two semicircles. This data set is characterized by flow pattern and density difference. Linear Segmentation is composed of two quasi-rectangulars. This data set is characterized by division of linear and density difference. Random noise is distributed in the data space of each data set.

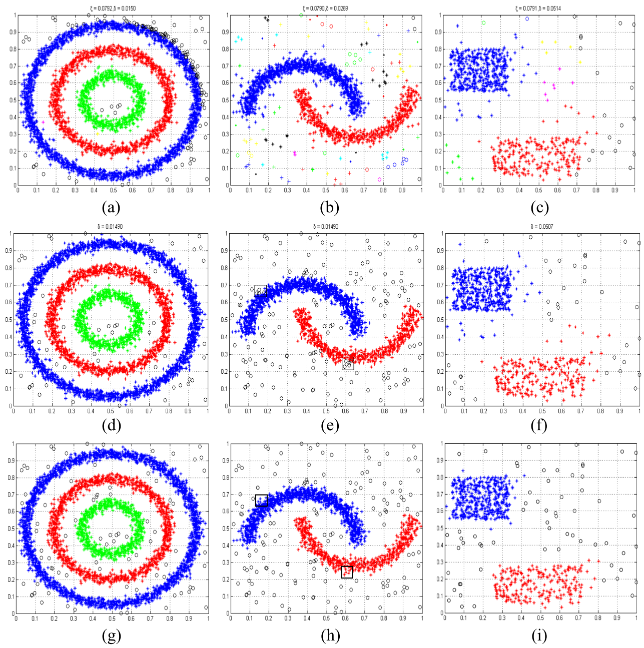

FIGURE 7. **Clustering results of our method, ABADNT, and original DENCLUE on different data sets.**

In the following experiment, the parameters h, ξ, δ should be set artificially in original DENCLUE in advance. For the following data sets from left to right, h is set as 5, ξ is set as 0.0792, 0.0790, 0.0791, and δ is set as 0.0150, 0.0269, 0.0514, respectively. In ABADNT, replaced by h and δ, ξ needs not be setčbut h is set as 5, and δ is set as 0.0149, 0.0149, 0.0507 from left to the right in the following experiment. In our method, ξ is replaced by h and δ. h, δ are determined according to the distribution of data in different data sets, so there are no parameters that should be set in advance.

The clustering results of the original DENCLUE are shown in upper part of Fig.7; the results of ABADNT are shown in middle part of Fig.7; the results of our method are shown in bottom part of Fig.7.

According to the clustering result, we can see that in the original DENCLUE, many noise points, especially those in the 2nd and 3rd data sets, are wrongly classified into clusters rather than being identified.

In ABADNT, the clustering result (shown in the Fig.7.(d)) of 1st data set is satisfying, but some data points (in the black rectangular) in 2nd data set (shown as the Fig.7.(e)) are wrongly identified as noise points, and some noise points (especially in the middle) in 3rd data set (shown as the Fig.7.(f)) are wrongly identified as data points.

In our method, although some noise points near to clusters cannot be identified, the performances in the 3 data sets are obviously better than those in the original DENCLUE and those in ABADNT. (Both data points in the black rectangular of Fig.7.(h) and noise points in the middle of Fig.7.(i) are identified correctly.)

The silhouette coefficients and the running time of clustering results in our method, the original DENCLUE and ABADNT are shown in Table 1.

TABLE 1. Silhouette coefficient of the clustering results with our method, ABADNT and original DENCLUE.

Methods	Datasets	Silhouette Coefficient	Running time (s)
Proposed Method	Three Circles	-0.1705	1.10237
	Two Moons	0.3414	1.02837
	Linear Segmentation	0.7172	0.98375
ABADNT	Three Circles	-0.1803	2.90338
	Two Moons	0.3302	2.73482
	Linear Segmentation	0.6974	2.55681
DENCLUE	Three Circles	-0.2863	4.49264
	Two Moons	0.3313	4.14733
	Linear Segmentation	0.6158	3.89947

We can see that for all of the data sets, the silhouette coefficients of results in our method are bigger than those in the original DENCLUE and ABADNT. The clustering result objectively reflects that the performance of our method is better. Additionally, the running times of Three Circles, Two Moons, and Linear Segmentation in our method are 1.10237s, 1.02837s, and 0.98375s, respectively. They are shorter than those in the original DENCLUE and ABADNT.

V. CONCLUSION AND FUTURE WORKS

The application range of the original DENCLUE is restricted, because many parameters need to be set manually in advance. In order to overcome this flaw, we present multi-parameter-setting based on data original distribution for DENCLUE optimization. In this algorithm, firstly, the value of smoothing parameter h is determined by data origional distribution. Then, for each data point, in the hill-climbing process of finding point's density attractor, step size δ will be adapted according to the density of data points. Lastly noise threshold ξ is replaced by the number of the data points in the set and step size δ. The experiment demonstrates that a better performance can be gotten through our method. But, if the noise points are close to or in the cluster, they will be classified into the cluster wrongly.

The following 2 aspects should be put emphasis in the future research.

(1) Whether the definition proposed in the paper for the value of h can be applicable to all data sets or not lacks rigorous mathematical proof.

(2) Which method we can adopt to identify the noise points close to the clusters.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[2] N. Padhy, P. Mishra, and R. Panigrahi, "The survey of data mining applications and feature scope," *Asian J. Comput. Sci. Inf. Technol.*, vol. 2, no. 4, pp. 43–58, 2014.

[3] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2013.

[4] A. Rodriguez and A. Laio, "Machine learning. Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, p. 1492, 2014.

[5] V. Sware and H. N. Bharathi, "Study of density based algorithms," *Int. J. Comput. Appl.*, vol. 69, no. 26, pp. 1–4, 2013.

[6] A. Hinneburg and H.-H. Gabriel, "DENCLUE 2.0: Fast clustering based on kernel density estimation," in *Proc. Int. Symp. Adv. Intell. Data Anal.*, vol. 4723. Mar. 2007, pp. 70–80.

[7] M. Suganya and S. Nagarajan, "Message passing in clusters using fuzzy density based clustering," *Indian J. Sci. Technol.*, vol. 8, no. 16, pp. 1–6, 2015.

[8] D.-X. Chang, X.-D. Zhang, C.-W. Zheng, and D.-M. Zhang, "A robust dynamic niching genetic algorithm with niche migration for automatic clustering problem," *Pattern Recognit.*, vol. 43, no. 4, pp. 1346–1360, 2010.

[9] C. Guo and Y. Zang, "Clustering algorithm based on density function and nichePSO," *J. Syst. Eng. Electron.*, vol. 23, no. 3, pp. 445–452, Jun. 2012.

[10] S. K. Sree, "SSM-DENCLUE : Enhanced approach for clustering of sequential data: Experiments and test cases," *Int. J. Comput. Appl.*, vol. 96, no. 6, pp. 7–13, 2014.

[11] L. Zhu, D. Li, F. He, Y. Yuan, and Y. Tong, "Fault diagnosis of belt weigher using the improved DENCLUE and SVM," *Harbin Gongye Daxue Xuebao/J. Harbin Inst. Technol.*, vol. 47, no. 7, pp. 122–128, 2015.

[12] J. Yan *et al.*, "Optimal clustering algorithm for crime spatial aggregation states analysis," *J. Tsinghua Univ.*, vol. 49, no. 2, pp. 176–178, 2009.

[13] K. Zhang *et al.*, "A novel algorithm based on avoid determining noise threshold in DENCLUE," in *Proceedings of Chinese Intelligent Systems Conference*. Singapore: Springer, 2016.

[14] H. Rehioui *et al.*, "DENCLUE-IM: A new approach for big data clustering," *Proc. Comput. Sci.*, vol. 83, no. 3, pp. 560–567, 2016.

[15] A. Idrissi, M. Abourezq, F. Zegrari, and H. Rehioui, "An improvement of DENCLUE algorithm for the data clustering," in *Proc. Int. Conf. Inf. Commun. Technol. Accessibility*, 2015, pp. 1–6.

[16] J. He and W. Pan, "A DENCLUE based approach to neuro-fuzzy system modeling," in *Proc. Int. Conf. Adv. Comput. Control*, 2010, pp. 42–46.

[17] K. Zhang *et al.*, "Pre-warning analysis and application in traceability systems for food production supply chains," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2500–2507, 2011.

[18] Z. Ke, C. Yi, L. Jianhuan, and F. Xiaohui, "Analysis of class group distinguishing based conceptual models for multiple fault diagnosis," in *Proc. Control Conf.*, Jul. 2015, pp. 6397–6402.

[19] K. Zhang, Y. Chai, and J. H. Liu, "A class of category discrimination based frameworks for multiple diagnosis," *Control Theory Appl.*, vol. 33, no. 2, pp. 154–163, 2015.

[20] Y. Xiaogao and Y. Xiaopeng, "On unsupervised clustering algorithm based on distance and density," *Comput. Appl. Softw.*, vol. 2, no. 2, pp. 1016–1021, 2010.

[21] A. Hinneburg and D. Keim, "An efficient approach to clustering in large multimedia databases with noise," in *Proc. 4th Int. Conf. Knowl. Discovery Data Mining*, 1998, pp. 58–65.

[22] C. C. Aggarwal, "Outlier analysis," in *Data Mining*. 2013, pp. 237–263.

[23] X. Wu and X. Zhu, "Mining with noise knowledge: Error-aware data mining," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 4, pp. 917–932, Jul. 2008.

[24] K. L. Zhou *et al.*, "On cluster validation," *Syst. Eng.-Theory Pract.*, vol. 34, no. 9, pp. 2417–2431, 2014.

[25] Y. Liu, Z. Li, H. Xiong, X. Gao, J. Wu, and S. Wu, "Understanding and enhancement of internal clustering validation measures," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 982–994, Jun. 2013.

YI LUO received the B.S. degree from Tianjin University in 2009, and the M.S. degree from Beijing Normal University in 2012. He is currently pursuing the Ph.D. degree with the School of Automation, Chongqing University.

His research interest covers community detection, clustering, and complex network.

KE ZHANG received the B.S. degree from Southwest Normal University in 2001, and the M.S. and Ph.D. degrees from Chongqing University in 2005 and 2010, respectively.

He is currently an Associate Professor with the School of Automation, Chongqing University. His research interest covers intelligent control, pattern classification, and fault diagnosis.

YI CHAI received the B.S. degree from the National University of Defense Technology in 1982, and the Ph.D. degree from Chongqing University in 2001.

He is currently a Professor with the School of Automation, Chongqing University. His research interest covers information processing, integration and control, and computer network and system control.

YINGZHI XIONG received the B.S. degree from the Chongqing University of Posts and Telecommunications in 2013, and the M.S. degree from Chongqing University in 2016.

She is currently with the China Automotive Engineering Research Institute Co., Ltd. Her research interest covers data analysis and automatic driving.