2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

K-Means Clustering with Modifying Firefly Algorithm

Masaki Takeuchi†, Thomas Ott‡, Haruna Matsushita§, Yoko Uwate† and Yoshifumi Nishio†

†Department of Electrical and Electronic Engineering, Tokushima University
2–1 Minami–Josanjima, Tokushima, 770–8506, Japan
‡Zurich University of Applied Sciences
Einsiedlerstrasse 31a, 8820 Waedenswil, Switzerland
§Department of Electronics and Information Engineering, Kagawa University
2217-20 Hayashi-cho, Takamatsu, Kagawa, 761–0396, Japan
Email: masaki@ee.tokushima-u.ac.jp, thomas.ott@zhaw.ch,
haruna@eng.kagawa-u.ac.jp, {uwate, nishio}@ee.tokushima-u.ac.jp

Abstract— In 2011, Senthilnath et al. proposed to utilize the Firefly Algorithm for K-means clustering. The algorithm has shown better results than the standard K-means algorithm or other combinations with bio-inspired optimization heuristics. In this study, we propose a further improvement of the method, based on an improved firefly algorithm. As a key aspect, the randomization parameter in our proposed algorithm is changed when the assignment does not change. We compare the standard K-means algorithm, K-means using the conventional Firefly Algorithm and our proposed algorithm on the basis of a simple data distribution. Numerical experiments show that our proposed algorithm is more efficient than the other algorithms.

1. Introduction

Clustering is a popular data analysis technique used for data analysis, image analysis, data mining and the other fields of science and engineering. The goal of clustering is to find homogeneous groups of data points in a data set. Each group is called a cluster and is characterized by the fact that objects that belong to the same group are more similar than objects that belong to different groups. The K-means algorithm is one of the most famous clustering methods. It is used if the number of clusters is known and the clusters tend to be spherical. The goal of the method is to find K cluster centers and assign each object to the closest cluster center such that the sum of the squared distances between the objects and the corresponding cluster centers is minimal. This means that the K-means clustering problem is an optimization problem.

Senthilnath et al. proposed an algorithm that used the firefly algorithm for K-means clustering (KMFA) [1]. Numerical experiments have indicated that this algorithm is more efficient algorithm than the standard algorithm or other optimization heuristics. The Firefly Algorithm (FA) has been proposed by Yang in 2007 and is based on the idealized behavior of the flashing characteristics of fireflies [2]. FA is an efficient optimization algorithm because it has a deterministic component and a random component. Almost all algorithms having only the deterministic component are local search algorithms, for which there is a risk of being trapped in a local optimum. However, the random component makes it possible to escape from such a local optimum.

In our previous study, we proposed a new clustering algorithm that combines K-means clustering and improved Firefly Algorithm (KMIFA). In our proposed algorithm, one parameter is changed when the assignment does not change [4]. We compared the conventional K-means algorithm, KMFA and our proposed algorithm KMIFA using a 2-dimensional toy data model. These experiments indicated that our algorithm is more efficient than the other algorithms. However, for 3-dimensional toy data model this algorithm cannot obtain a better results than other algorithms. We improved the transition rule of one parameter [5]. Our proposed algorithm has new two parameters. In the previous studies, we carried out computer simulations with fixed parameters. Therefore, in this study, we simulate a various patterns of this two parameters and intestate their effects.

2. The Conventional Methods

In this section, we explain the conventional K-means algorithm and the Firefly Algorithm (FA).

2.1. K-means algorithm

The objective function of K-means clustering is defined by

$$J = \sum_{k=1}^{K} \sum_{n=1}^{N} b_{kn} |\boldsymbol{c}_k - \boldsymbol{o}_n|^2, \tag{1}$$

where K is the number of cluster centers, N is the number of objects, $\boldsymbol{c_k}$ is the position vector of cluster center k and $\boldsymbol{o_n}$ is the position vector of object n. Each object is assigned to its nearest center. Hence $b_{kn} = 1$ if object n is assigned

Algorithm 1 The conventional Firefly Algorithm

```
Objective function f(x), x = (x_1, ..., x_d)^T
Initialize a population of fireflies x_i (i = 1, 2, ..., n)
while t < MaxGeneration do
  for i = 1 to n, all n fireflies do
    for j = 1 to n, all n fireflies do
      Light intensity I_i at x_i is determined by f(x_i)
      if I_i > I_j then
        Move firefly i towards j in all d dimensions
      end if
      Attractiveness varies with distance r via exp[-γr]
      Evaluate new solutions and update light intensity
    end for j
  end for i
  Rank the fireflies and find the current best
end while
Postprocess results and visualization
```

to center k and $b_{kn} = 0$ otherwise:

$$b_{kn} = \begin{cases} 1, & object\ n\ is\ assigned\ to\ center\ k \\ 0, & otherwise. \end{cases} \tag{2}$$

This optimization problem is solved using the K-means algorithm which is composed of the following four steps:

1. Initialize all cluster centers and objects: The number of cluster centers and all objects are predefined. All cluster centers are randomly initialized in the search space.

2. Assignments: Each object is assigned to only the closest cluster center.

3. Calculates cluster centers: The places of each cluster center move to the mean of each group object.

4. Iterate steps 2 and 3 until the assignments no longer change.

2.2. The Conventional Firefly Algorithm (FA)

FA has been developed by Yang and it was based on the idealized behavior of the flashing characteristics of fireflies [2]. The conventional FA idealizes these flashing characteristics using the following three rules:

- All fireflies are unisex so that one firefly is attracted to other fireflies regardless of their sex.

- Attractiveness is proportional to brightness; thus, for any two flashing fireflies, the less brighter one will move towards the brighter one. Both the attractiveness and brightness stared above decrease as their distance increases. If no one is brighter than a particular firefly, it moves randomly.

- The brightness or light intensity of a firefly is affected or determined by the landscape of the objective function to be optimized.

The attractiveness of a firefly β is defined by

$$\beta = (\beta_0 - \beta_{min})e^{-\gamma r_{ij}^2} + \beta_{min}, \tag{3}$$

$$\gamma = \frac{1}{\sqrt{L}}, \tag{4}$$

$$L = \frac{|X_{max} - X_{min}|}{2}, \tag{5}$$

where γ is the light absorption coefficient, β_{min} is the minimum value of β, β_0 is the attractiveness at $r_{ij} = 0$, and r_{ij} is the Euclidian distance between any two fireflies i and j at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. L means the average scale for the problem. The movement of the firefly i is attracted to another more attractive firefly j, and is determined by

$$\boldsymbol{x}_i = \boldsymbol{x}_i + \beta(\boldsymbol{x}_j - \boldsymbol{x}_i) + \alpha\boldsymbol{\epsilon}_i, \tag{6}$$

$$\boldsymbol{\epsilon}_i = (random() - 0.5)L, \tag{7}$$

where $\boldsymbol{x}_i$ is the position vector of firefly i, $random()$ is a uniform random number distributed in [0, 1] and $\alpha(t)$ is the randomization parameter. The parameter $\alpha(t)$ is defined by

$$\alpha(t) = \alpha(0)\left(\frac{10^{-4}}{0.9}\right)^{t/t_{max}}, \tag{8}$$

where t is the number of iteration.

Algorithm 1 shows the pseudo code of the conventional FA for minimum optimization problems.

3. K-Means Clustering with FA (KMFA)

For KMFA, the position vector $\boldsymbol{x}_i$ of a firefly i corresponds to $(\boldsymbol{c}_1, \boldsymbol{c}_2, ..., \boldsymbol{c}_K)$. That is, each firefly contains the positions of all cluster centers. The attractiveness of each firefly is defined by the objective function (Eq. (7)). Numerical experiments have indicated that this algorithm is more efficient than the K-means algorithm and other algorithms for typical benchmark data sets [1].

4. K-means Clustering using the Improved FA (KMIFA)

The K-means algorithm and KMFA sometimes converge to a local minimum. Therefore, the purpose of this study is to remove this disadvantage. In our proposed algorithm, each firefly has its own value of $\alpha(t)$:

$$\alpha(t) = \lambda_i\left(\frac{10^{-4}}{0.9}\right)^{t/t_{max}}, \tag{9}$$

where λ is a new parameter. We set all value of λ to the same certain value λ_0 when initializing the population of fireflies and define the minimum value of λ is 0. We do not

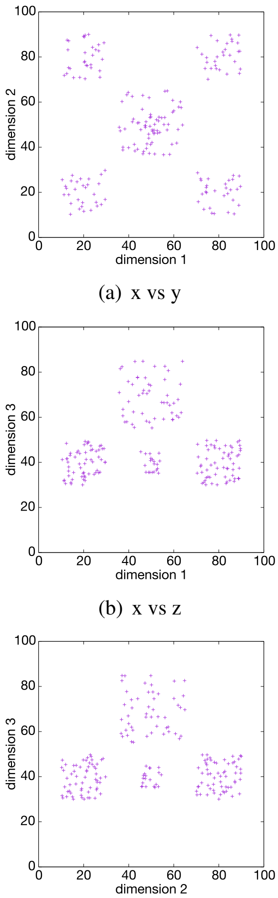
Table 1: Information about data toy model used

cluter	ideal center	object number	ball of radius
1	$(50, 50, 70)$	50	15
2	$(20, 20, 40)$	30	10
3	$(20, 80, 40)$	30	10
4	$(80, 20, 40)$	30	10
5	$(80, 80, 40)$	30	10
6	$(50, 50, 40)$	20	5

define the maximum value of λ, which means λ could increase to infinity. The value of λ changes if the assignment changes or not.

$$\lambda_i^{new} = \begin{cases} \lambda_i - \theta_1, & \textit{the assignment doesn't change} \\ \lambda_i + \theta_2, & \textit{the assignment changes} \end{cases} \tag{10}$$

where θ_1 and θ_2 are a predefined parameter. In the case of a firefly i, if the assignment of all objects does not change, the value of λ_i decreases. On the other hand, if the assignment of all objects changes, the value of λ_i increases. In the case of $\lambda >> 0$, a firefly moves with a relatively strong random influence. This makes the firefly easier to escape from a local minimum. In the case of $\lambda = 0.0$, a firefly does not move randomly, which leads to a faster convergence. Therefore, the concept of our proposed algorithm is at the beginning of the search, fireflies easily escape from local optima. Then, as the number of iteration increases, fireflies tend to converge.



(a) x vs y

(b) x vs z

(c) y vs z

Figure 1: Data toy model used.

5. Numerical Experiments

We compare the conventional K-means algorithm, KMFA and two transition patterns our KMIFA using a simple data toy model. Information about that model is summarized in Tab. 1 and that model is depicted in Fig. 1. The number of dimensions is 3, the range of each dimension is $[0, 100]$, the number of clusters is 6 and the number of total objects is 190. The data objects were generated randomly around the ideal centers within each ball of radius. We used all the same data set in each numerical experiment. Each numerical experiment was run 500 times and we compared the success rate of each algorithm, where the success rate is defined as the fraction of objects that are assigned to the correct cluster:

$$Success\ Rate[\%] = \frac{Success\ Times[times]}{500} \times 100 \tag{11}$$

Figure 2 shows the numerical experiment results in the case of $\theta_1 = 0.1$ and $\theta_2 = 0.1$. We assume that our proposed algorithm is more efficient algorithm than the other two algorithms. For our proposed algorithm, the success rates are almost same as those of KMFA when λ_0 is more than 1.5. As λ_0 decreasing from 1.5, the success rates of our proposed algorithm are gradually increasing. On the other hand, As λ_0 decreasing from 1.5, the success rates of KMFA remind flat to 0.5, then, those rapidly decrease.

Figure 3 shows the transition of λ when λ_0 is 0.0. Each line means the value of λ of each firefly. The transition is like a mountain. Until the number of iterations is about 30, λ of all fireflies increase. From the number of iterations is about 60, λ of all fireflies decrease.

Next, we focus on the rule of changing λ. Figure 4 shows the numerical experiment results in the case θ_1 is fixed at 0.1 and θ_2 is 0 and 0.1. We assume the algorithm having only decrease operation cannot obtain good result.

Next, we change the value of θ_1 and θ_2. First, we fix the value of θ_1 at 0.1 and change θ_2 from 0.1 to 0.3 (see Fig. 5). Figure 5 shows the value of θ_2 is suitable for our proposed algorithm. The graph of our proposed algorithms gradually increase with the same slope. However, as the value of θ_2 increasing, the success rate decreases.

Then, we fix the value of θ_2 at 0.1 and change θ_1 from 0.1 to 0.3 (see Fig. 6). We assume that the success rate does not depend on the value of θ_2.

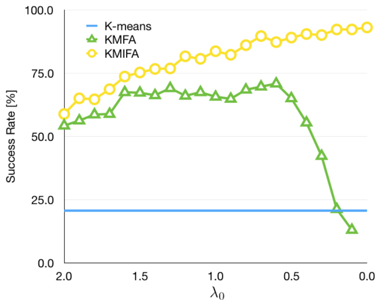


Figure 2: Numerical experiment result.

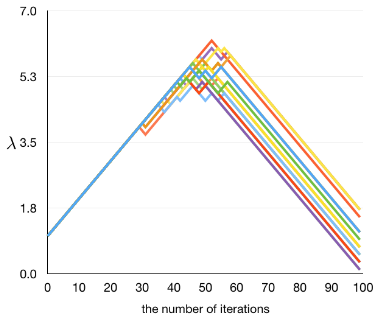


Figure 3: The tradition of $\lambda_0 = 0.0$.

6. Conclusion

In this study, we have proposed a new clustering algorithm that utilizes an improved firefly algorithm for K-means clustering. Our algorithm is based on the idea that the randomization parameter is changed when the assignment changes or not. In our proposed algorithm, at the beginning of the search, all fireflies move with a relatively strong random influence. Hence they can more easily escape from a local minimum. As the number of iterations increases, the firefly tend to converge. Numerical experiments have indicated that our proposed algorithm is more efficient than the other algorithms.

The study is based on a relatively simple toy data set. In our future work, we will study more complex problems. In addition, we will study more various transition.

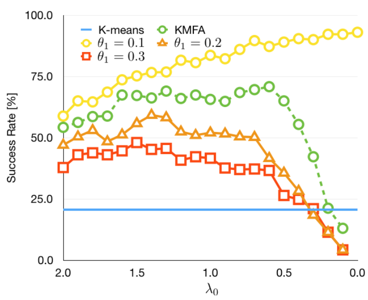


Figure 4: Numerical experiment result only decrease operation.

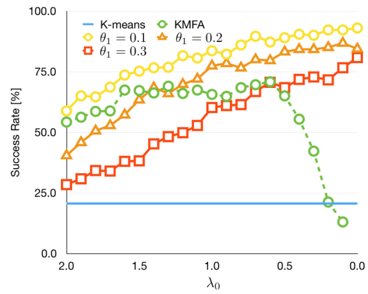


Figure 5: Numerical experiment result changing θ_2.

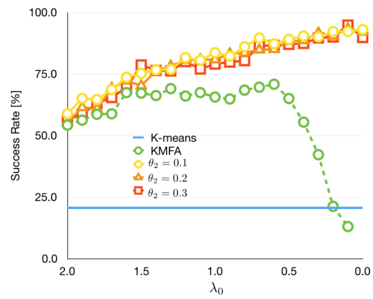


Figure 6: Numerical experiment result changing θ_1.

Acknowledgment

This work was partly supported by JSPS Grant-in-Aid for Scientific Research 16K06357.

References

[1] J. Senthilnath, S.N. Omkar and V. Mani, *"Clustering using Firefly Algorithm: Performance Study"*, Swarm and Evolutionary Computation 1, pp. 164–171, 2011.

[2] X.S. Yang, *Nature–Inspired Metaheuristic Algorithms Second Edition*, Luniver Press, 2010.

[3] J. Kennedy and R. Eberhart, *"Particle Swarm Optimization"*, Proceedings of the International Conference on Neural Networks, pp. 1942–1948, 1995.

[4] M. Takeuchi, T. Ott, H. Matsushita, Y. Uwate and Y. Nishio, *"K-Means Algorithm using Improved Firefly Algorithm"*, Proceedings of 2017 RISP International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'17), pp. 225–228, 2017.

[5] M. Takeuchi, T. Ott, H. Matsushita, Y. Uwate and Y. Nishio , *"Investigation of K-Means Clustering Used Firefly Algorithm Modified Random Component"*, Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics, 2017 (will appear).