2017 RISP International Workshop on Nonlinear Circuits,
Communications and Signal Processing (NCSP'17)
Guam, USA, February 28th to March 3rd, 2017

K-means Algorithm using Improved Firefly Algorithm

Masaki Takeuchi[1], Thomas Ott[2], Haruna Matsushita[3], Yoko Uwate[1] and Yoshifumi Nishio[1]

1 : Department of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima, 770–8506, Japan
E-mail: {masaki, uwate, nishio}@ee.tokushima-u.ac.jp

2 : Zurich University of Applied Sciences
Einsiedlerstrasse 31a, 8820 Waedenswil, Switzerland
E-mail: thomas.ott@zhaw.ch

3 : Department of Electronics and Information Engineering, Kagawa University
2217-20 Hayashi-cho, Takamatsu, Kagawa, 761-0396, Japan
E-mail: haruna@eng.kagawa-u.ac.jp

Abstract

In 2011, K-means algorithm combined Firefly Algorithm has been proposed by Mr. Senthilnath. This clustering algorithm has obtained better results than K-means algorithm and the other algorithms combined bio-inspired algorithm. In this study, we propose a new clustering algorithm; K-means algorithm combined improved Firefly Algorithm. One parameter of our proposed algorithm is changed when the assignment does not change. We compare K-means algorithm, K-means algorithm combined Firefly Algorithm and our proposed algorithm using simple model. Numerical experiments show our proposed algorithm is more efficient algorithm than the other algorithms.

1. Introduction

A clustering is a popular data analysis technique and is 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 each object can be defined as a relations between objects belong in the same group are stronger than a relations between objects belong in the different group. K-menas algorithm is one of the most famous clustering method and is used if the number of clusters is known and the clusters tend to be spherical. In this algorithm, we can find K cluster centers (each object is assigned to the closest cluster center). In other words, this algorithm can find the minimum sum of distances between the object and the cluster center it belongs. This means that the clustering is one of a optimization problems.

Mr. Senthilnath et. al. [1] had proposed a new clustering algorithm combined K-means algorithm and Firefly Algorithm and numerical experiments have indicated that it is more efficient algorithm than the others. Firefly Algorithm (FA) has been proposed by Mr. Yang in 2007 and is based on the idealized behavior of the flashing characteristics of firefly [2]. FA is a 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 at local optima. However, the random component makes it possible to escape from such local optima.

In this paper, we propose a new clustering algorithm combined K-means algorithm and Improved FA. In our proposed algorithm, one parameter is changed when the assignment does not change. We compare the conventional K-means algorithm, K-means algorithm combined the conventional FA and our proposed algorithm using simple model. Numerical experiments indicate that our proposed algorithm is more efficient algorithm than the other algorithms.

2. The Conventional Methods

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

2.1 K-means algorithm

K-menas clustering is one of the most famous clustering method and is used if the number of clusters is known and the clusters tend to be spherical. The goal of this clustering is to find cluster centers and each object is assigned to the closest cluster center. Therefore this clustering is a optimization problem because we need to find the minimum sum of Euclidean distances between the cluster centers and the objects.

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
```

This clustering is solved using K-means algorithm and this algorithm is composed of the following four steps:

- Initialize all cluster centers and objects. The number of cluster centers and all objects are predefined. All cluster centers generate randomly in search space.
- Assignments. Each object is assigned to only the closest cluster center.
- Calculates cluster centers. The places of each cluster center move to mean of each group object.
- Iterate assignment and calculate steps until the assignments no longer change.

The objective function of clustering is defined by

$$J = \sum_{k=1}^{K} \sum_{n=1}^{N} |\boldsymbol{c_k} - \boldsymbol{o_n}|^2, \tag{1}$$

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.

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]. Many active researchers have paid attention to FA [3, 4] 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.

Attractiveness of firefly β is defined by

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

$$\gamma = \frac{1}{\sqrt{L}}, \tag{3}$$

$$L = \frac{|X_{max} - X_{min}|}{2}, \tag{4}$$

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 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

$$\boldsymbol{x_i} = \boldsymbol{x_i} + \beta(\boldsymbol{x_j} - \boldsymbol{x_i}) + \alpha\boldsymbol{\epsilon_i}, \tag{5}$$

$$\boldsymbol{\epsilon_i} = (random - 0.5)L, \tag{6}$$

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

$$\alpha(t) = \alpha(0)\left(\frac{10^{-4}}{0.9}\right)^{t/t_{max}}, \tag{7}$$

where t is the number of iteration.

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

3. K-means algorithm combined FA

The clustering algorithm combined K-means algorithm and the conventional FA has been proposed in 2011 by Mr. Senthilnath et. al [1]. In this algorithm, the position vector of firefly i $\boldsymbol{x_i}$ is $(\boldsymbol{c_1}, \boldsymbol{c_2}, ..., \boldsymbol{c_K})$, that is, each firefly has the places of all cluster centers. The attractiveness of each firefly is defined by objective function (Eq. (7)). Numerical experiments have indicated that this algorithm is more efficient algorithm than K-means algorithm and other algorithms using well known typical benchmark data sets.

Algorithm 2 shows pseudo code of K-means algorithm combined FA.

Algorithm 2 K-means algorithm combined FA

Initialize a population of fireflies $\boldsymbol{x_i} (i = 1, 2, ..., N_f)$
Initialize a cluster centers $\boldsymbol{c_k} (k = 1, 2, ..., N_c)$
Initialize a objects $\boldsymbol{o_n} (n = 1, 2, ..., N_o)$
while $t < MaxGeneration$ **do**
  Each object is assigned the closest cluster center
  calculate $J_i (i = 1, 2, ..., Nf)$
  Light intensity I_i is determined by J_i
  if $I_i > I_j$ **then**
    Move firefly i towards j
  end if
  Find the current best
end while
Postprocess results and visualization

4. K-means algorithm combined Improved FA

K-means algorithm sometimes converges to a local minimum from a starting position of the search. K-means algorithm combined FA also has this disadvantage. Therefore, our purpose of this study is to remove this disadvantage. In our proposed algorithm, each firefly has their own value of $\alpha(0)$ and we set all values of $\alpha(0)$ to 1.0 at the same time of initializing a population of fireflies. In the case of a firefly i, if the assignment of all objects does not change, the value of $\alpha(0)$ of firefly i decreases. We define the minimum value of $\alpha(0)$ is 0.0.

$$\alpha(0)_i^{new} = \begin{cases} \alpha(0)_i - V & , \textit{the assignment changes} \\ \alpha(0)_i & , \textit{otherwise} \end{cases} \tag{8}$$

The parameter V is predefined certain value. In the case of $\alpha(0) = 1.0$, a firefly moves relatively randomly. This makes the firefly easier to escape a local minimum. In the case of $\alpha(0) = 0.0$, a firefly does not move randomly, which makes the firefly easier to converge.

5. Numerical Experiments

We compare the conventional K-means algorithm, K-menas algorithm combined the conventional FA and our proposed algorithm using simple model we defined.

We explain the simple model (see Fig. 1). We set the number of dimensions is 2, each dimension range is [0, 100], the number of clusters is 6 and the number of objects is 120 (20 objects are belong to one cluster). Each ideal cluster center is predefined at (30, 30), (70, 70), (10, 80), (90, 20), (50, 90) and (70, 70). Each object is re-generated around ideal cluster center by simulate, the distances from the objects to the allocated cluster centers are within 10. We set $\beta_0 = 1.0$, $\beta_{min} = 0.1$ and $N_f = 20$. Each numerical experiment is run 300 and we compare success rate of each algorithm. The

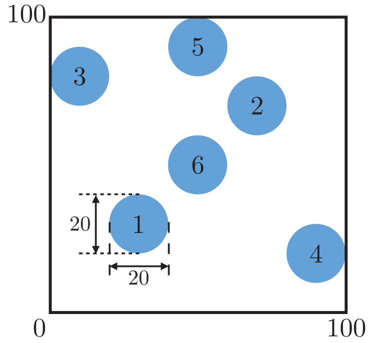

Figure 1: simple modal we used

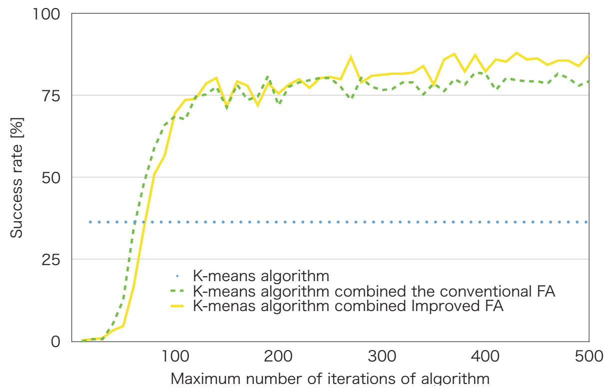

Figure 2: Comparison of three algorithms using simple modal

success we defined is that all objects are assigned at correct cluster.

Figure 2 shows numerical experimental results when the parameter $V = 0.1$ We assume that our proposed algorithm is more efficient algorithm than the other two algorithms. Our proposed algorithm and K-means algorithm combined the conventional FA are more efficient algorithm when the maximum number of iterations of algorithm is more than about 70. Our proposed algorithm is more efficient algorithm when the maximum number of iterations of algorithm is more than about 300.

Next, we focus on the parameter V. We set the value of V to 0.1, 0.2 and 0.3 and compare success rate (see Fig. 3). As increasing the parameter V, success rate goes up quickly until the maximum number of iterations of algorithm is 100. On the other hand, when the maximum number of iterations of algorithm is from 500 to 700, we hard to determine which value is better than others according to this graph. Therefore,

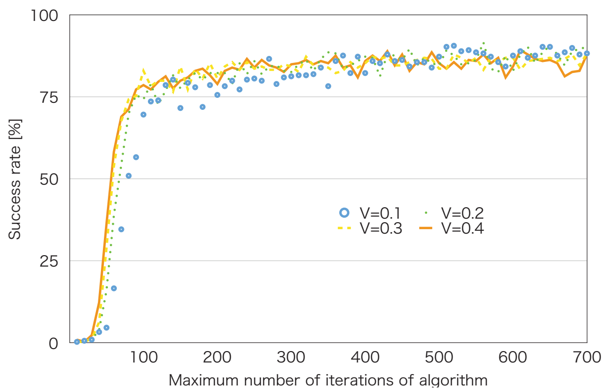

Figure 3: Comparison of the parameter V

Figure 4: Comparison of the right time to decrease the parameter $\alpha(0)$

Table 1: The average, maximum and minimum values of success rate from 500 to 700

value of V	average	maximum	minimum
0.1	88.4	90.7	84.3
0.2	87.2	91.7	82.7
0.3	86.1	88.3	83.3
0.4	85.4	89.3	81.0

we summarize a average, miximum and minimum values of success rate in Tab. 1. In the case of average and minimum values, algorithm set $V = 0.1$ performs better. In the case of maximum value, algorithm set $V = 0.2$ performs better. We assume that $V = 0.1$ is the most suitable value.

Finally, we focus on the right time to decrease the parameter $\alpha(0)$. From here, we decrease the parameter $\alpha(0)$ when the new assignment is the same as only the last assignment. We change the right time to decrease the parameter $\alpha(0)$ when the new assignment is the same as the last two, three four or five assignment (see Fig 4). We assume that the right time to decrease the parameter $\alpha(0)$ is when the new assignment is the same as the last assignment. Success rates until the maximum number of iterations of algorithm is about 100 are almost same. On the other hand, when maximum number of iterations of algorithm is from 300, the case of only last assignment obtains better result than the others.

6. Conclusion

In this study, we have proposed a new clustering algorithm; K-means algorithm combined improved Firefly Algorithm. In our proposed algorithm, one parameter is changed when the assignment does not change. In our proposed algorithm, at the beginning of the search, all fireflies move relatively randomly. This mean firefly easy to escape from a local minimum. As increasing iteration of algorithm, firefly tend to converge. Numerical experiments have indicated that our proposed algorithm is more efficient algorithm than other two algorithms.

In future work, we try to improve our proposed algorithm and apply more complex problems.

Acknowledgment

This work was partly supported by JSPS Grant-in-Aid for Challenging Exploratory Research 26540127.

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] S. Lukasik and S. Zak, *"Firefly Algorithm for Continuous Constrained Optimization Tasks"*, Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, Vol. 5796 of the Series Lecture Notes in Computer Science, pp. 97–106, 2009.

[4] H. Matsushita, *"Firefly Algorithm with Dynamically Changing Connections"*, Proceedings of International Symposium on Nonlinear Theory and its Application (NOLTA'14), pp. 906–909, 2014.

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