Investigation of K-means Algorithm Using an Improved Firefly Algorithm

M. Takeuchi[1], **T. Ott**[2], **H. Matsushita**[3], **Y. Uwate**[1] **and Y. Nishio**[1]

1: Tokushima Univ., 2: Zurich Univ. of Applied Sciences, 3: Kagawa Univ.

1: {masaki, uwate, nishio}@ee.tokushima-u.ac.jp, 2: thomas.ott@zhaw.ch,

3: haruna@eng.kagawa-u.ac.jp

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 belonging to the same group are more similar than objects belonging 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. [1] proposed an algorithm that used the firefly algorithm for K-means clustering (KMFA). 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 [2–4]. 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 this paper, we propose a new clustering algorithm that combines K-means clustering and improved firefly algorithm (KMIFA). In our proposed algorithm, each firefly has its own value of $\alpha(t)$:

$$\alpha(t)_i = \lambda_i \left(\frac{10^{-4}}{0.9} \right)^{t/t_{max}}. \tag{1}$$

In the case of firefly i, if the assignment of all objects does not change, the value of λ_i decreases. We set all initial values of λ to 1.0 when initializing the population of fireflies and define the minimum value of λ is 0.

$$\lambda_i = \begin{cases} \lambda_i^{old} - V & \text{, the assignment does not change} \\ \lambda_i^{old} & \text{, otherwise} \end{cases} \tag{2}$$

The parameter V is a predefined value. At the beginning of the search, all fireflies move with a relatively strong random influence. Hence, they can more easily escape from local optima. As the number of iterations increases, the firefly tends to converge. We compare the conventional K-means algorithm, KMFA and our proposed algorithm KMIFA using several data models that have several spherical clusters. These experiments indicate that our algorithm is more efficient than the other algorithms.

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.

The preliminary version of this study is presented at NCSP2017.