Investigation of a New Clustering Method used Modified Firefly Algorithm for K-Means Clustering

Masaki Takeuchi
Dept. of Electrical
and Electronic Eng.,
Tokushima University
2-1 Minami-Josanjima,
Tokushima, 770–8506, Japan
masaki@ee.tokushima-u.ac.jp

Thomas Ott
Zurich University
of Applied Sciences
thomas.ott@zhaw.ch

Haruna Matsushita
Dept. of Electronics
and Info. Eng.,
Kagawa University
2217-20 Hayashi-cho,
Takamatsu, Kagawa,
761–0396, Japan
haruna@eng.kagawa-u.ac.jp

Yoko Uwate and Yoshifumi Nishio
Dept. of Electrical
and Electronic Eng.,
Tokushima University
2-1 Minami-Josanjima,
Tokushima, 770–8506, Japan
{ uwate , nishio }@ee.tokushima-u.ac.jp

SUMMARY

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. One of the most famous clustering methods is the K-means algorithm. It is used if the number of clusters is known and the clusters tend to be spherical. The goal of this 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], [3], [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 study, we propose a new clustering method that combines K-means clustering and improved firefly algorithm (KMIFA). We modify the random component of FA to suit our concept. In KMIFA, 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, the previous Senthilnath's study 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.