Bee Colony Optimization for Clustering

Alexey Oleynik, Sergey Subbotin and Andrey Oleynik

Abstract – **The method of clustering based on the bee colony optimization is proposed. The main advantage of the developed method is the estimation of the number of cluster in a given dataset where no previous knowledge of the number of clusters is available.**

Keywords – **Agent, Swarm Intelligence, Clustering, Sample. Bee Colony Optimization.**

I. INTRODUCTION

One of the newest branches of optimization methods is Swarm Intelligence based on multi-agent approach and inspired by modeling of the behaviour of living organisms. There is many methods of Swarm Intelligence: Ant Colony Optimization, Bee Colony Optimization, Particle Swarm Optimization and others.

The method based on bee behaviour modeling is developed for clustering in this work. There is many methods based on bee behaviour modeling. But all of them are used for numerical optimization. The main concept of them is following: there is three types of agents (scouts, employed foragers and unemployed foragers), scouts find new solutions in search space randomly and then by usage of waggle dance modeling they recruit unemployed foragers which become employed foragers; employed foragers investigate more detail solutions found by scouts.

II. CLUSTERING BASED ON BEE COLONY OPTIMIZATION

The proposed multi-agent method for clustering is based on bee behaviour modeling. When bees explore their natural environment they pollinate it. And then as result of pollinating areas of similar objects arise. In the proposed method these areas are considered as clusters.

The developed multi-agent method for clustering consists of the following stages.

1. **Initializing**. Search space is created. Agents and objects (instances) of base sample are randomly scattered in created search space.

2. **Agents moving with choice of objects for transporting on search space**.

All agents are moving in search space and looking for objects for transporting. Agent takes copy of object (like pollen of flower) and then moves with it on search space.

3. **Agents moving and redundancy (duplication) of chosen objects**.

All agents are moving and checking cells for redundancy of taken and transported objects. When agent moves to cell with some objects it makes decision based on objects in the cell. It may be one of the following cases.

Alexey Oleynik, Sergey Subbotin, Andrey Oleynik – Department of software, Zaporozhye National Technical University, Zhukovskiy Str, 64, Zaporozhye, 69063, UKRAINE,
E-mail: lexAO1@gmail.com, subbotin@zntu.edu.ua, olejnik@zntu.edu.ua

1. There is just one object in cell. Agent j decides on the redundancy of transported object o^j randomly with 50% probability.

2. There is more than one object in cell.

2.1. There is the object with worse conditions in cell than conditions for transported object. In this case agent j takes $o_{worst}^{l,p}$ for transporting: $o^j = o_{worst}^{l,p}$.

2.2. In the investigated cell conditions for transported objects are better than in source cell. In this case agent j decides the redundancy of transported object o^j: $o^{l,p} = \{o^{l,p}, o^j\}$.

3. Otherwise. Agent searches some other cell with objects.

Moving of agents (stage 2 and 3) can be carried out some times for better search space exploring and better transporting of chosen objects.

4. **Natural selection**: leaving of objects in cells with best conditions for them.

Each agent can leave object in more than one position. That's why after transporting each object can be located in more than one cell and selection is needed: each object must stay in just one cell.

Thus, each object stays in cell with best conditions.

5. **Merging of objects in cells**. After some execution of stages 2–4 merging of objects in cells is carried out. It is necessary for better aggregation of objects in clusters. All objects in each cells are merged to one new object.

And on the following iterations agents work with new merged objects.

Developed clustering method based on bee colony optimization was realized in Matlab 7.0. To validate the approach described in the present work developed software is used for clustering of the sample of synthetic data.

III. CONCLUSION

In this work, a new method for clustering was described. This method is based on multi-agent approach and is inspired by modeling of bee behaviour.

Foundations of the proposed method were defined. It was noted that the main stages of the developed method are: initializing, agent moving (for object choice and for object transporting), natural selection and merging. The main advantages of the developed method are: a priori information about shape of clusters is not necessary and definition of cluster numbers is not needed.

The created program was applied to benchmark solving. It was used on synthetic data (generated by Gaussian deviation). The results obtained by the usage of developed program are found to be correct.

That's why we can conclude that the developed multi-agent method for clustering can be an efficient tool for clustering in real-world problems.