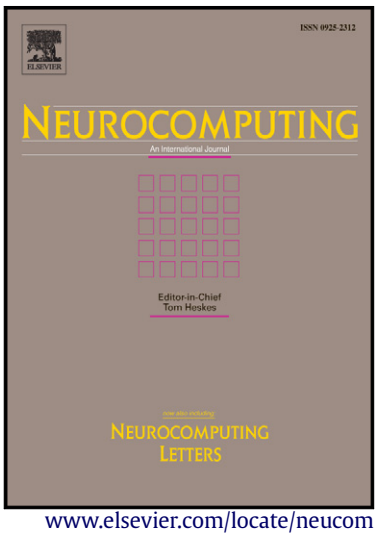


www.elsevier.com/locate/neucom

# An improved bee colony optimization algorithm with an application to document clustering

Rana Forsati, Andisheh Keikha, Mehrnoush Shmasfard

PII: S0925-2312(15)00212-X
DOI: http://dx.doi.org/10.1016/j.neucom.2015.02.048
Reference: NEUCOM15179

To appear in: *Neurocomputing*

Received date: 23 September 2013
Revised date: 20 November 2014
Accepted date: 5 February 2015

Cite this article as: Rana Forsati, Andisheh Keikha, Mehrnoush Shmasfard, An improved bee colony optimization algorithm with an application to document clustering, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2015.02.048

# An Improved Bee Colony Optimization Algorithm with an Application to Document Clustering

Rana Forsati[a], Andisheh Keikha[b], Mehrnoush Shmasfard[a]

[a] *Natural Language Processing (NLP) Research Lab., Faculty of Electrical and Computer Engineering, Shahid Beheshti University, G. C., Tehran, Iran*
[b] *Department of Computer Science and Engineering, Ryerson University*

{r_forsati, m_shams}@sbu.ac.ir, andisheh.keikha@ryerson.ca, mahdavim@cse.msu.edu

**Abstract**

The bee colony optimization (BCO) algorithm is proved to be one of the fast, robust and efficient global search heuristics in tackling different practical problems. Considering BCO algorithm in this paper, we utilize it for the data clustering, a fundamental problem that frequently arises in many applications. However, we discovered some obstacles in directly applying the ancient BCO to address the clustering problem and managed to change some basic behaviors of this swarm algorithm. In particular, we present an improved bee colony optimization algorithm, dubbed IBCO, by introducing *cloning* and *fairness* concepts into the BCO algorithm and make it more efficient for data clustering. These features give BCO very powerful and balanced exploration and exploitation capabilities to effectively guide the search process toward the proximity of the high quality solutions. In particular, the cloning feature allows it to take advantage of experiences gained from previous generations when generating new solutions. The problem of getting stuck in a local optima still laid bare in the proposed improved version. As a result, to overcome the shortage of this swarm algorithm in searching locally, we hybridize it with the $k$-means algorithm to take advantage of fine tuning power of the widely used $k$-means algorithm which demonstrates good result in local searches. We propose four different hybridized algorithms based on IBCO and $k$-means algorithms and investigate the clustering results and convergence behavior of them. We empirically demonstrate that our hybrid algorithms alleviate the problem of sticking in a local solution even for large and high dimensional data sets such as document clustering. The results show that proposed algorithms are robust enough to be used in many applications compared to $k$-means and other recently proposed evolutionary based clustering algorithms including genetic, particle swarm optimization, ant colony, and bee based algorithms.

*Keywords:* Swarm Intelligence, Bee Colony Optimization, Document Clustering



## 1. Introduction

Clustering is one of the crucial unsupervised learning techniques for dealing with massive amounts of heterogeneous information. The aim of clustering is to group a set of data objects