# Clustering Web service Data Using Artificial Bee Colony (ABC) Algorithm

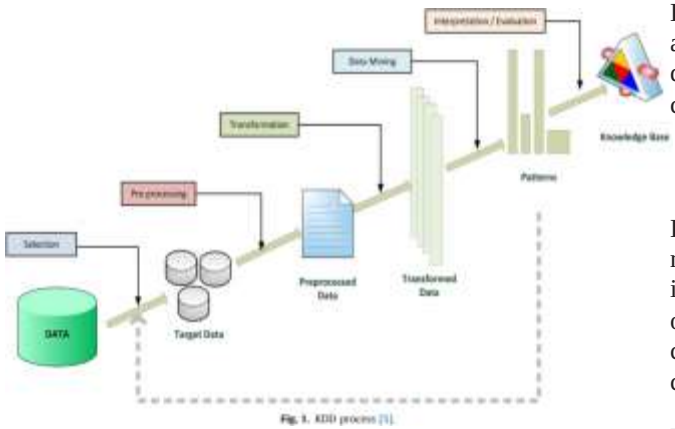
**B.Bhavana[#1], D.Varun Kumar[#2], K.Satya[#3,] M.Madan Mohan Malavya[#4].**
[#1, 2,, 3, 4] *Students CSE Department, Lendi institute of engineering and technology, Vizianagaram, Andhra Pradesh, India.*
[1]bhavanabalasa117@gmail.com, [2]dvarun94@gmail.com, [3]satyakoratana@gmail.com, [4]madan.lendi@gmail.com.

***Abstract*— Data clustering is one of the most popular techniques in data mining. Clustering high-dimensional data is the cluster analysis of data which have anywhere from a few dozen to many thousands of dimensions. Such high-dimensional data spaces are often encountered in areas. There are many Web service companies which provide their services to society contain the data i.e., high dimensional data. Different optimization techniques have been applied to investigate the optimal solution for clustering problems but the accuracy is low when applied to high dimensional data. Swarm intelligence (SI) is one such optimization technique whose algorithms have successfully been demonstrated as solutions for different data clustering domains with high dimensions. Artificial bee colony algorithm (ABC) is an optimization algorithm obtained from SI. It based on the intelligent foraging behaviour of honey bee swarm. We analyze the algorithm which had given the optimised solution.**

***Keywords*-- High-dimensional data, Web service, Swarm Intelligence, Artificial bee colony.**

## I. INTRODUCTION

Extracting information from large data becomes difficult. On one hand, the increase in the amount of data adds to the possibility of the data possessing more information but on the other hand, it decreases the speed of pattern extraction. Knowledge Discovery and Data mining (KDD) has made possible the discovery of such useful and hidden patterns in the data.




KDD is the process of automatically searching large volumes of data for previously unknown, potentially interesting and informative patterns. KDD techniques are mainly influenced by modern information exploration techniques; however, they also rely on traditional computational techniques from statistics, information retrieval, machine learning and pattern recognition.

Data clustering, one of the most important techniques in data mining aims to group unlabeled data into different groups on the basis of similarities and dissimilarities between the data elements.

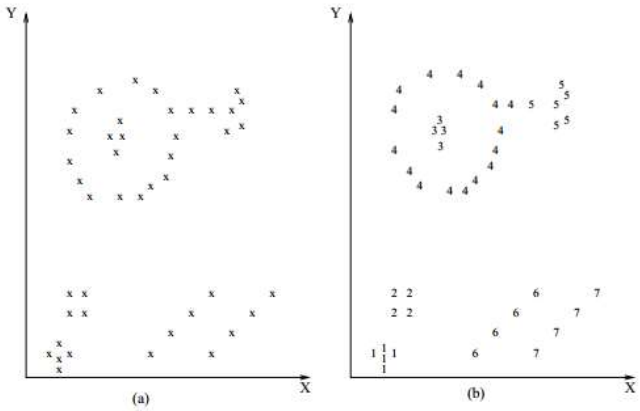



Fig. 1. Data clustering.

A typical clustering process involves feature selection, selection of a similarity measure, grouping of data, and assessment of the output. The process can be performed in a supervised, semi supervised, or unsupervised manner. Different algorithms have been proposed that take into account the nature of the data, the quantity of the data, and other input parameters in order to cluster the data..

## II. EXISTING SYSTEM

Data clustering has received a lot of attention from researchers of various data mining domains. This has resulted in a number of approaches being suggested to address one or other aspects of the data clustering problem. Two of the most commonly used clustering approaches are partition-based clustering and hierarchy-based clustering.

Partition-based clustering divides the data into different groups based on similarities and dissimilarities among the data elements. Similarity measures differ from application to application, but the most common measures are distance-based, pattern-based, and density-based similarity measures. In distance-based similarity measures, a distance function is

used to calculate the relative position of a data element inside the cluster by comparing it with the center of the cluster i.e. the centroid. K-means clustering technique is one of the foundations of the partitional approaches.

Hierarchical clustering provides a sequence of nested partitions of the dataset in the form of a hierarchy. It divides the data into a nested tree structure where the levels of the tree show similarity or dissimilarity among the clusters at different levels. A hierarchical approach is either agglomerative or divisive. A divisive approach is based on the splitting of one large cluster into sub clusters. In the agglomerative approach, the clustering process starts with every data element in an individual cluster. The visualization of hierarchical clustering can be shown using a dendrogram.

**DEMERITS:**

Limitations occurred when clustering is applied to high dimensional data are

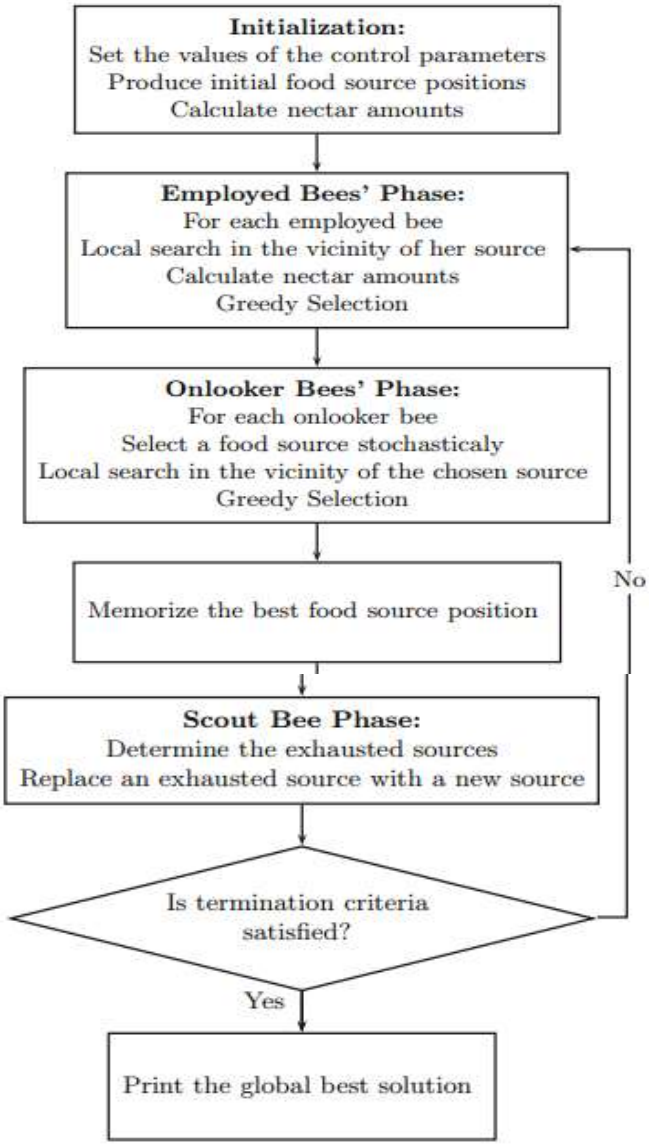
1. Poor accuracy for high dimensional data as it cannot handle it.
2. Efficiency is low.
3. Apart from their advantages, both techniques have deficiencies in terms of shape and structure of clusters, exact number of clusters, clustering configuration and degeneracy.

To tackle these problems, optimization-based techniques have been investigated for data clustering.

## III. PROPOSED SYSTEM

Artificial Intelligence (AI) is one of the oldest and best known research fields. Swarm intelligence in evolutionary computation exploits the behaviour information obtained from self-organizing agents (swarm) without supervision (DE). The term swarm is used for an aggregation of animals like fishes, birds and insects such as ants, termites and bees performing collective behaviour. SI is defined as the collective behaviour of decentralized and self-organized swarms.

The approaches which are based on bee colony, called artificial bee colony (ABC). The collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers and unemployed foragers, and the model defines two leading modes of the behaviour: the recruitment to a rich nectar source and the abandonment of a poor source.




**1** Flowchart of the ABC algorithm

The colony of artificial bees in ABC contains three groups of bees: *employed bees* associated with specific food source, *onlooker bees* watching the dance of employed bees within the hive to choose a food source, and *scout bees* searching for food sources randomly. Both onlookers and scouts are also called unemployed bees. Initially, all food source positions are discovered by scout bees. Thereafter, the nectar of food sources are exploited by employed bees and onlooker bees, and this continual exploitation will ultimately cause them to become exhausted. Then, the employed bee which was exploiting the exhausted food source becomes a scout bee in search of further food sources once again. In other words, the employed bee whose food source has been exhausted becomes a scout bee. In ABC, the position of a food source represents a possible solution to the problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. In the basic form, the number of

employed bees is equal to the number of food sources (solutions) since each employed bee is associated with one and only one food source.

**MERITS:**

1. Accuracy is maintained as each bee has their correspond functionalities.
2. Performance is high as the task is shared by them.

## REFERENCES

[1]. Banharnsakun A, Achalakul T, Sirinaovakul B (2010a) Abc-gsx: A hybrid method for solving the traveling salesman problem. In: 2010 second world congress on nature and biologically inspired computing (NaBIC), pp 7–12

[2]. Cobanli S, Ozturk A, Guvenc U, Tosun S (2010) Active power loss minimization in electric power systems through artificial bee colony algorithm. Int Rev Electr Eng-IREE 5(5, Part b):2217–2223

[3]. Dorigo M, Blum C (2005) Ant colony optimization theory: a survey. Theor Compute Sci 344(2-3):243–278

[4]. Garro BA, Sossa H, Vazquez RA (2011)Artificial neural network synthesis by means of artificial bee colony (abc) algorithm. In: 2011 IEEE congress on evolutionary computation (CEC), pp 331–338

[5]. Horng MH, Jiang TW (2010) Multilevel image thresholding selection using the artificial bee colony algorithm.In: Wang F, Deng H, Gao Y, Lei J (eds) Artificial intelligence and computational intelligence. Lecture notes in computer science, vol 6320. Springer, Berlin, pp 318–325

[6]. Irani R, Nasimi R (2011) Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling. J Pet Sci Eng 78(1):6–12

[7]. Ji P, Wu Y (2011) an improved artificial bee colony algorithm for the capacitated vehicle routing problem with time-dependent travel times. In: Tenth international symposium on operations research and its applications (ISORA 2011), pp 75–82

[8]. Kang F, Li J, Xu Q (2009a) Hybrid simplex artificial bee colony algorithm and its application in material dynamic parameter back analysis of concrete dams. J Hydraul Eng 40(6):736–742