Swarm Intelligence for feature selection: A review of literature and reflection on future challenges

Nandini Nayar[1], Sachin Ahuja[2], Shaily Jain[1]

[1] Department of Computer Science and Engineering, Chitkara University Himachal Pradesh, India
[2] Chitkara UniversityResearch &Innovation Network, Chitkara University, Punjab, India

Abstract. Feature subset selection is considered to be a significant task in data mining. Researchers are working to develop optimal solutions with higher order of computational efficiency. This paper presents a review on problems encountered during the process of feature selection and how Swarm Intelligence has been used for extraction of optimal set of features. It also gives a concise overview of various Swarm Intelligence algorithms like Particle Swarm Optimization, Ant Colony Optimization, Bacteria Foraging Algorithms, Bees Algorithm, BAT algorithms and the various hybrid approaches that have been discovered using these approaches. Finally, some considerations of the research gaps and the future scope of Swarm Intelligence have been included.

1 Introduction

The dataset with a large number of features (attributes) is known as high-dimensional dataset. The accuracy of a model can be enhanced by using a wisely selected subset of features, rather than using every available feature. Researchers have proved that feature subsets give superior results as compared to entire set of features for the same algorithm.

1.1 Feature Selection

Databases that contain a huge amount of data, with high-dimensionality are exceptionally common. Feature selection may be referred as the process of selecting the optimal subset of features based on certain criterion [1].
Feature selection and variable selection are essential for knowledge discovery from huge amount of data. Traditional variable selection methods (e.g. Cp, AIC and BIC) involve a combinatorial optimization problem that is NP-hard. Their computational time increases with the dimensionality. Due to this expensive computational cost, these traditional procedures become fairly infeasible for high-dimensional data analysis. Thus, more innovative variable selection procedures are essential to deal with high-dimensionality [2].Feature selection has a vital role in data pre-processing techniques that are used for data mining [3].