International Journal of Software Science and Computational Intelligence
Volume 11 • Issue 2 • April-June 2019

A Novel Chaotic Northern Bald Ibis Optimization Algorithm for Solving Different Cluster Problems [ICCICC18 #155]

Ravi Kumar Saidala, Acharya Nagarjuna University, Guntur, India
https://orcid.org/0000-0002-3658-6651

Nagaraju Devarakonda, Lakireddy Bali Reddy College of Engineering, Mylavaram, India

ABSTRACT

This article proposes a new optimal data clustering method for finding optimal clusters of data by incorporating chaotic maps into the standard NOA. NOA, a newly developed optimization technique, has been shown to be efficient in generating optimal results with lowest solution cost. The incorporation of chaotic maps into metaheuristics enables algorithms to diversify the solution space into two phases: explore and exploit more. To make the NOA more efficient and avoid premature convergence, chaotic maps are incorporated in this work, termed as CNOAs. Ten different chaotic maps are incorporated individually into standard NOA for testing the optimization performance. The CNOA is first benchmarked on 23 standard functions. Secondly, testing was done on the numerical complexity of the new clustering method which utilizes CNOA, by solving 10 UCI data cluster problems and 4 web document cluster problems. The comparisons have been made with the help of obtaining statistical and graphical results. The superiority of the proposed optimal clustering algorithm is evident from the simulations and comparisons.

KEYWORDS

Chaotic Maps, Cognitive Algorithms, Data Clustering Problems, Nature-Inspired Metaheuristics, NOA

INTRODUCTION

Generating and sharing of the magnitude of data via public administrations, business, scientific research, numerous industrial and non-profit sectors has increased immeasurably. These data include textual content (i.e. unstructured, semi structured as well as structured (Hashimi et al., 2015), to multimedia content (e.g. audio, images, videos) on a variety of platforms (e.g. sensor networks, system-to-system communications, cyber-physical systems, social media websites and Internet of Things) (Witten et al., 2016). Due to the incessant growth in generating and sharing of data, new and efficient techniques are needed for accessing, discovering the hidden knowledge and sharing the same from various domains (Larose et al., 2014). Human investigation for knowledge extraction of this huge data is a tiresome task and it was found that the obtained results are no longer accurate. The classical algorithms are inaccurate in interpreting and extracting hidden knowledge. So, new

DOI: 10.4018/IJSSCI.2019040101

and advanced technologies are needed to come into existence to understand the knowledge extraction process automatically and summarize the meaningful information as per the application requirements (Thuraisingham, 2014). Therefore, it is an obligation to design clever and efficient techniques to analyze this massive data. Since 1990's when data mining techniques have appeared in database family, it is broadly used to extract hidden knowledge and pattern from enormous data sets (Han, 2011). This extraction uses two different techniques, namely supervised and unsupervised techniques (Brownlee, 2016). Clustering is the most used unsupervised and popular data analysis technique in data mining for extracting the hidden knowledge of data by partitioning it into clusters or groups. The ultimate purpose of clustering is to generate the clusters of similar data objects by classifying the unlabeled input data. By doing this, the similarity is to be minimized between the objects of each cluster while the similarity is also to be maximized between objects of other clusters. Hierarchical and Partitional clustering are the two primary categories of the developed numerous clustering algorithms (Jain, 2010). The first category algorithms seek to build a tree structure of cluster in the absence of prior knowledge about the count of initial clusters. In the second category, an initial cluster centroid is assigned. The k-means partitional clustering technique is the widely used and the most prevalent algorithm. This technique effectively groups extensive datasets based on the best runtime. In spite of the fact that the k-means algorithm is quicker than numerous other algorithms, it experiences two note-worthy issues, i.e. exhibiting high sensitivity in the initialization phase and local optima at a low convergence rate (Jain, 2010; Kantardzic, 2011). It has been noticed from the literature (Alam et al., 2014; Esmin et al., 2015; José-García, & Gómez-Flores, 2016; Nanda & Panda, 2014; Saidala & Devarakonda, 2018a) that conjoining the nature–inspired optimization algorithms with standard data clustering techniques will result in accurate solutions. It also enables to overcome the drawbacks found in the standard data clustering techniques.

Optimization is an action of obtaining the minimization or maximization of a function to the available resources and is subject to certain constraints. Authors Brownlee (2011) and Yang (2008) quoted that, by concerning the nature of optimization, optimization algorithms are categorized into deterministic and stochastic intelligent algorithms. Deterministic algorithms are gradient restricted algorithms which move rigorously towards the optimal solution. These kinds of algorithms produce the predefined output and exhibit known same amount of time or memory or resources every time to the given same input. Unlike the deterministic algorithms, stochastic algorithms are gradient free algorithms. The contradictory thing of stochastic algorithms to the deterministic algorithm is random steps are to be taken by stochastic algorithms to reach the optima. Stochastic algorithms are further classified into two groups, i.e., Heuristics and Metaheuristics (Yang et al., 2012). First group algorithms work by trial-and-error method, whereas the second group works by having some prior knowledge about the random search. Several heuristic algorithms have been used in optimization field; for instance, Hill Climbing (HC) (Tsamardinos, 2006), Simulated Annealing (SA) (Aarts & Korst, 1988; Szu & Hartley, 1987) and Bat Swarm Optimization (BSO) (Jordehi, 2015). In meta-heuristic algorithms, optimization process starts with a random solution. From that initial solution Metaheuristics explore and exploits the solution space randomly with a certain probability (Nanda & Panda, 2014).

The last two decades witnessed the development of Computational Intelligence (CI) techniques and their applications in various design and engineering fields. Many researchers working in various domains have consequently turned to CI in hope of discovering optimal solutions to a plethora of complex problems (Primeau et al., 2018). A subfield of CI, i.e., nature-inspired metaheuristics is robust and efficient in solving NP-hard problems. The Metaheuristics is becoming more and more popular due to their robustness and efficiency in dealing with real-world *NP*-hard problems (Bouarara et al., 2015; Feng et al., 2018; Gheraibia et al., 2015; Nouaouria et al., 2014; Saidala & Devarakonda, 2017a; Saidala & Devarakonda, 2017b; Schor & Kinsner, 2011). Data clustering is one of the most used and significant data analysis concepts. The ultimate purpose of clustering is to generate the clusters of similar data objects by classifying the unlabeled input data where the similarity should

Detection of Cyber Crime Based on Facial Pattern Enhancement Using Machine Learning and Image Processing Techniques