Alexandria Engineering Journal (2017) xxx, xxx–xxx

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

Multi kernel and dynamic fractional lion optimization algorithm for data clustering

Satish Chander [a,*], P. Vijaya [a], Praveen Dhyani [b]

[a] Waljat College of Applied Sciences, P.O. Box 197, P.C. 124, Rusayl, Muscat, Oman
[b] Banasthali University, Jaipur Campus, India

Received 27 August 2016; revised 8 November 2016; accepted 15 December 2016

KEYWORDS

Data clustering;
Multi kernel function;
Fractional lion optimization;
Directive operative searching strategy;
Clustering accuracy

Abstract Clustering is the technique used to partition the homogenous data, where the data are grouped together. In order to improve the clustering accuracy, the adaptive dynamic directive operative fractional lion algorithm is proposed using multi kernel function. Also, we intend to develop a new mathematical function for fitness evaluation. We utilize multi kernels, such as Gaussian, tangential, rational quadratic and Inverse multiquadratic to design the new fitness function. Consequently, the WLI fuzzy clustering mechanism is employed in this paper to determine the distance measurement based on new fitness function, named Multi kernel WLI (MKWLI). Then, we design a novel algorithm with the aid of dynamic directive operative searching strategy and adaptive fractional lion algorithm, termed Adaptive Dynamic Directive Operative Fractional Lion (ADDOFL) algorithm. Initially in this proposed algorithm, the solutions are generated based on the fractional lion algorithm. It also exploits the new MKWLI fitness function to evaluate the optimal value. Finally, the updation of female lion is performed through dynamic directive operative searching algorithm. Thus, the proposed ADDOFL algorithm is used to find out the optimal cluster center iteratively. The simulation results are validated and performance is analyzed using metrics such as clustering accuracy, Jaccard coefficient and rand coefficient. The outcome of the proposed algorithm attains the clustering accuracy of 89.6% for both Iris and Wine databases which ensures the better clustering performance.

1. Introduction

Nowadays, the database is the rapid development paradigm, widely used in several information technology applications such as production, marketing, business and so on [1]. Due to such enormous amount of data, the data mining is one of the significant approaches to extort the useful information [2–4]. Data mining is the novel concept used to cater the knowledge among various fields such as investigation of medical records, analysis of client transactions and market surveying applications [5]. The clustering and classification algorithm are the two main prerequisites in data mining to extract the information. Hence, the data clustering becomes the facilitating and challenging task to generate the clusters, where the data are grouped together [6–9]. The two types of clustering

* Corresponding author.
E-mail address: dimrisatish@gmail.com (S. Chander).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

http://dx.doi.org/10.1016/j.aej.2016.12.013

mechanism are utilized to perform the data clustering, (i) hierarchical clustering and (ii) partitional clustering [10]. In hierarchical clustering, the data objects are clustered in the hierarchical manner either in the form of tree and dendrogram. On the contrary, the non-overlapping clusters are obtained by the partitional clustering mechanism [11]. In real-time applications, this algorithm has the tendency to deal with noise and outliers, scalability, interoperability, usability and so on. Some of the commonly used clustering algorithms are K-means clustering, Fuzzy c-means algorithm, density based clustering and so on [12–14].

In fuzzy clustering, the data clustering is performed based on the membership function. On the other hand in the traditional algorithm, distance measure is the major concern to evaluate the similarity measure between two data objects. Albeit the clustering algorithm is developed, the major issue incurs in data clustering is NP hard problem since each data has high degree of similarity with other parts of data. To resolve this issue, the researcher exploits optimization algorithm to find out the optimal clusters for data clustering [15]. Some of the optimization algorithms are Genetic algorithm [16], Particle Swarm Optimization (PSO) [17], Artificial Bee Colony (ABC) [18], Differential Evolution Algorithm [19], firefly optimization [20], etc. Among these, the genetic algorithm is initially developed and employed for the data clustering mechanism. Then, the fitness function is the noteworthy aspect to select the optimal cluster in search space of the solution [21]. In most of the optimization based clustering algorithms, the foremost problem is convergence and local optimal solution trap. Furthermore, the computational burden also becomes the cumbersome for data clustering. In order to surmount the drawbacks of aforementioned optimization algorithm, we propose novel adaptive dynamic directive operative fractional lion (ADDOFL) optimization based clustering algorithm assists with the multi-kernel function in this paper. Thus, the proposed clustering algorithm ensures to improve the clustering accuracy compared to existing optimization based data clustering algorithms.

The two main contributions of this paper are as follows:

Adaptive Dynamic Directive Operative Fractional Lion Optimization algorithm: ADDOFL optimization algorithm is the main contribution of this paper. The proposed algorithm comprises of dynamic decision operative searching strategy and adaptive fractional lion algorithm. In ADDOFL, the female lion updation is performed in the fractional lion algorithm using dynamic directive operative searching strategy.

Fitness function: The four kernel functions such as Gaussian kernel, tangential kernel, rational quadratic kernel and inverse multiquadratic kernel are utilized. Then, the WLI fuzzy clustering exploits this new fitness function to evaluate the distance measurement between data object and cluster center. Using the proposed fitness function, the best optimal solution is chosen for data clustering.

This paper is structured as follows: Section 2 discusses the data clustering mechanism from eight research papers and also advantages and disadvantages are given. The problem specification and challenges behind the approach are presented in Section 3. Section 4 briefly explains the proposed algorithm of adaptive dynamic directive operative fractional lion algorithm. The simulation results and performance are analyzed in Section 5. Finally, this paper is concluded in Section 6.

2. Literature review

Huang et al. [23] demonstrated the extend K-means clustering algorithm. The extension was made by combining the cluster distribution. The integration of intra cluster compactness and inter cluster separations was using Wk-means, and AWA. The advantage of their approach was robustness since balancing parameters were not required in cluster distribution. But, the shortcoming is that the division by zero problem. Parker et al. [24] designed the data clustering algorithm by altering the Fuzzy C-means algorithm. They intended to manipulate the FCM algorithm with the aid of Geometric Progressive. Thus, it became easy to handle small range of error, but it mitigates the quality. A fuzzy rough set based clustering algorithm was introduced by Maji [25]. According to rough set theory, the data clustering was performed. The key advantage is to remove the redundant or irrelevant data using quantitative measure. This measure was incorporated with sample categories to assign the class labels.

Bandyopadhyay et al. [26] developed the multi-objective optimization technique for data clustering. They utilized the multiple objective functions to generate the clusters. Using cluster validity index, the optimization algorithm was used to provide the better scalability measure. But, the main disadvantage of their method is critical to identify the non-convex clusters. Motamedi et al. [27] explained a kernel algorithm for data clustering. The kernel function in algorithm was used to increase the dimension of feature space to perform the clustering mechanism and also mitigate the requirement of approximation function. It was used to reduce the complexity issue but more sensitive to changes in corpus. Binu et al. [28] presented multi-kernel function based data clustering method using fuzzy c-means and cuckoo optimization algorithm. The developed clustering process was used to determine the optimal solution with the aid of variance measure among data points. Through the multiple objective based kernel function, the optimal centroid was selected using cuckoo search algorithm. The hybrid algorithm resulted in clustering with higher 96% of data accuracy.

Naldi and Campello [29] described an evolutionary K-means algorithm based data clustering. Since K-means is said to be simple, ease implementation and modifiable, adaptable to variety of contexts and applications domains, it was more vulnerable to distance reflection for centroid selection. They intent to incorporate both evolutionary algorithm and k-means algorithm used to conquer the drawback of k-means algorithm. The data partitions from k-means algorithm were also solved with the aid of adopted fast evolutionary algorithm. Data clustering using PSO and Bee colony optimization algorithm was developed by Dhote et al. [30]. The Bee algorithm was interpreted with the particle swarm optimization algorithm where they easily determined the optimal centroid for cluster generation. Due to integration of Bee colony, the local optimum problem of PSO was resolved. Due to parameter change of every iteration of optimization algorithm, the clustering performance gets degraded. The review outcome reveals the state-of-the-art clustering algorithm for data clustering and the critics on them.

Please cite this article in press as: S. Chander et al., Multi kernel and dynamic fractional lion optimization algorithm for data clustering, Alexandria Eng. J. (2017), http://dx.doi.org/10.1016/j.aej.2016.12.013