# Chapter 36
# On Combining Nature–Inspired Algorithms for Data Clustering

**Hanan Ahmed**
*Ain Shams University, Egypt*

**Safwat Hamad**
*Ain Shams University, Egypt*

**Howida A. Shedeed**
*Ain Shams University, Egypt*

**Mohamed F. Tolba**
*Ain Shams University, Egypt*

## ABSTRACT

*This chapter proposed different hybrid clustering methods based on combining particle swarm optimization (PSO), gravitational search algorithm (GSA) and free parameters central force optimization (CFO) with each other and with the k-means algorithm. The proposed methods were applied on 5 real datasets from the university of California, Irvine (UCI) machine learning repository. Comparative analysis was done in terms of three measures; the sum of intra cluster distances, the running time and the distances between the clusters centroids. The initial population for the used algorithms were enhanced to minimize the sum of intra cluster distances. Experimental results show that, increasing the number of iterations doesn't have a noticeable impact on the sum of intra cluster distances while it has a negative impact on the running time. K-means combined with GSA (KM-GSA), PSO combined with GSA (PSO-GSA) gave the best performance according to the sum of intra cluster distances while K-means combined with PSO (KM-PSO) and KM-GSA were the best in terms of the running time. Finally, KM-GSA and GSA have the best performance.*

## INTRODUCTION

Recently, with the advent of large memory systems and computer networks, the amount of stored data grows very quickly. These data contain useful but hidden information that may be extracted for various purposes. Organizing data into sensible groupings or clusters is one of the most fundamental modes of understanding and learning. Clustering is defined as the unsupervised classification of patterns (observations, data items or feature vectors) into groups (clusters). Clustering algorithms are used in many fields and applications such as: Document clustering and information retrieval applications, where the modified differential evolution algorithm was used in (Abraham, Das, & Konar, 2006, Hassanien, 2016),

hybrid harmony search was used for clustering web documents in (Mahdavi, Chehreghani, Abolhassani, & Forsati, 2008) and dynamic hierarchical compact and dynamic hierarchical star were used in (Gil-Gara & Pons-Porrata, 2010) for document clustering; Data compaction applications, where clustering based vector quantization algorithms and set partitioning in hierarchical trees were used in (Yang, Wu, Wang, & Jiao, 2010) for image compression; Anomaly detection applications, where crisp and fuzzy approaches based on the cosine similarity principle were used in (Friedmana, Lastb, Makover, & Kandel, 2007), Genetic algorithm was used in (Kerr, Ruskin, Crane, & Doolan, 2008) and dynamic agglomerative was used in (Liang & Wang, 2007) to cluster gene expression data ; Medicine applications, where spatially constrained kernel clustering was used in (Liao, Lin, & Li, 2008); Computer vision applications, where genetic algorithm, C-means, competitive learning and hierarchicalcompetitive learningwere used in (Scheunders, 19997) and single point iterative weighted fuzzy C-means was usedin (Fan, Han, & Wang, 2009); Construction management applications, where constrained k-prototypes were used in (Cheng & Leu, 2009); Marketing and customer analysis applications, where mixed integer programming was used in (Saglam, Salman, Sayın, & Türkay, 2006) and many other fields where the clustering algorithms were used in the primary task for understanding the nature and the structure of data or used in the pre-processing or post-processing phase for higher level tasks.

Local optimum of an optimization problem is a solution that is optimal (either maximal or minimal) within a neighbouring set of candidate solutions. There is a solution which is the optimal solution among all possible solutions (global optima). From our study for different clustering algorithms, it has been found that the local optimum is the problem of some algorithms and to overcome this problem, search based algorithms (optimization algorithms which are nature based) were used to get optimum centroids(Jain, Murty, & Flynn, 1999, Yamany et al 2015a,b, Tharwat et al. 2015, Hassanien, Aboul Ella & Alamry,E., 2015).

In this paper nature inspired algorithms are combined with K-means and with each other in various orders and a comparative study among these algorithms and their combinations are introduced. Comparative Analysis is done in terms of three measures; the sum of intra cluster distances, the running time and the distances between the clusters centroids. The effect on the performance and the run time for the different number of iterations is also studied.

The rest of the paper is organized as follows: Section 2 provides a brief explanation for data clustering types. The related work will be introduced in Section 3. A theoretical comparative study among the three basic algorithms PSO, GSA and CFO will be introduced in Section 4. The proposed experiments will be explained in Section 5 and the performance will be tested using the famous UCI repository datasets in Section 6. Finally, the conclusion of the work will be presented in section 7.

## BACKGROUND

## Central Force Optimization (CFO) Algorithm

Central force optimization (CFO) (Formato R., 2007) is a *deterministic* nature-inspired optimization that performs multi-dimensional global search and optimization. It is based on the gravitational kinematics. It searches the space for the extrema of the objective function to be maximized and models "probes" that "fly" through the decision space by analogy to masses moving under the influence of gravity. At the beginning, the initial probe positions are computed in a deterministic manner.

## Related Content

### The Formal Design Models of Tree Architectures and Behaviors

Yingxu Wangand Xinming Tan (2011). *International Journal of Software Science and Computational Intelligence (pp. 84-108).*

www.irma-international.org/article/formal-design-models-tree-architectures/64181

### Abstract Retrieval over Wikipedia Articles Using Neural Network

Falah Hassan Ali Al-akashi (2019). *International Journal of Software Science and Computational Intelligence (pp. 26-43).*

www.irma-international.org/article/abstract-retrieval-over-wikipedia-articles-using-neural-network/236150

### A Robust P2P Information Sharing System and its Application to Communication Support in Natural Disasters

Takuma Oide, Akiko Takahashi, Atsushi Takedaand Takuo Suganuma (2013). *International Journal of Software Science and Computational Intelligence (pp. 20-39).*

www.irma-international.org/article/a-robust-p2p-information-sharing-system-and-its-application-to-communication-support-in-natural-disasters/108928

### P3 Process for Object-Relational Data Migration to NoSQL Document-Oriented Datastore

Aicha Aggouneand Mohamed Sofiane Namoune (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-20).*

www.irma-international.org/article/p3-process-for-object-relational-data-migration-to-nosql-document-oriented-datastore/309994

### Neural Network (NN) Based Route Weight Computation for Bi-Directional Traffic Management System

Shamim Akhter, Rahatur Rahmanand Ashfaqul Islam (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications (pp. 750-765).*

www.irma-international.org/chapter/neural-network-nn-based-route-weight-computation-for-bi-directional-traffic-management-system/237903