Pertanika J. Sci. & Technol. 24 (2): 295 - 306 (2016)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Data Clustering using Differential Search Algorithm

Vijay Kumar[1]*, Jitender Kumar Chhabra[2] and Dinesh Kumar[3]

[1]Department of Computer Science and Engineering, Thapar University, Patiala, Punjab, India-147001

[2]Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India-136119

[3]Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Haryana, India-125001

ABSTRACT

The main challenges of clustering techniques are to tune the initial cluster centres and to avoid the solution being trapped in the local optima. In this paper, a new metaheuristic algorithm, Differential Search (DS), is used to solve these problems. The DS explores the search space of the given dataset to find the near-optimal cluster centres. The cluster centre-based encoding scheme is used to evolve the cluster centres. The proposed DS-based clustering technique is tested over four real-life datasets. The performance of DS-based clustering is compared with four recently developed metaheuristic techniques. The computational results are encouraging and demonstrate that the DS-based clustering provides better values in terms of precision, recall and G-Measure.

Keywords: Data clustering, differential search algorithm, metaheuristic

INTRODUCTION

Data clustering is an unsupervised technique that partitions a set of data points into a number of clusters such that data points in the same cluster are similar to each other according to some criteria. It has been found applicable in various applications such as engineering (artificial intelligence, pattern recognition, mechanical engineering), computer sciences (image segmentation, web mining), medical sciences (biology, microbiology), earth sciences (geology, remote sensing), social sciences (sociology, psychology) and economics (Everitt, 1993; Abraham et al., 2008). Data clustering algorithms are broadly classified into two main categories: hierarchical and partitional (Leung et al., 2000). The hierarchical clustering algorithms provide a tree structure output that represents the nested grouping of the elements of a dataset (Frigui & Krishnapuram, 1999).

Article history:
Received: 8 July 2015
Accepted: 5 October 2015

E-mail addresses:
vijaykumarchahar@gmail.com (Vijay Kumar),
jitenderchhabra@gmail.com (Jitender Kumar Chhabra),
dinesh_chutani@yahoo.com (Dinesh Kumar)
*Corresponding Author

ISSN: 0128-7680

These do not require *a priori* knowledge about the number of clusters present in the dataset. The main drawback of hierarchical clustering algorithms is that these may fail to separate overlapping clusters (Jain et al., 1999).

On the other hand, partitional clustering algorithms divide the dataset into a number of groups based upon certain clustering criteria. The clustering criterion directly affects the nature of clusters formed. The advantages of partitional clustering algorithms are the disadvantages of the hierarchical algorithms and vice versa. A large number of partitional clustering algorithms have been reported in the literature (Jain et al., 1999; Xu & Wunsch, 2009). The focus of this paper is on the partitional clustering algorithm. It is also worth mentioning that the hierarchical clustering algorithms are unable to differentiate the overlapping clusters and are computationally expensive. The well-known partitional clustering technique is K-Means. The performance of the K-Means clustering algorithm is heavily influenced by the number of clusters specified and the random choice of initial cluster centers. To solve these problems, a lot of metaheuristic techniques are available for use, such as genetic algorithm (GA), ant colony optimisation (ACO), particle swarm optimisation (PSO), gravitational search algorithm (GSA).

Recently, Civicioglu (2012) developed a novel metaheuristic algorithm, namely, the differential search algorithm (DS), which mimics the Brownian-like random-walk movement used for migration of organisms. DS is preferred over other metaheuristic techniques as it is easy to implement and has fewer control parameters. Due to these advantages of the DS, it is employed on data clustering. The novel approach is proposed, in which DS is used to generate the optimal cluster centres.

This paper aimed to explore the applicability of differential search approach to the development of clustering technique. It includes a general overview of data clustering with emphasis on recently developed metaheuristic-based clustering techniques followed by the proposed Differential Search-based clustering technique. The performance evaluation has been done on real-life datasets.

BACKGROUND

This section describes the related concept of clustering and related works.

Basic Concepts of Clustering

The mathematical formulation of partitional clustering technique is described (Abraham et al., 2008). Consider a dataset X that consists of n data points, $X=\{x_1,x_2,....,x_n\}$. Each data point is described by d features, where $x_j=\left(x_{j1},x_{j2},....,x_{jd}\right)$ is a vector representing the j^{th} data point and x_{ji} represents the i^{th} feature of x_j. As we know, the main objective of any clustering technique is to partition the dataset into a number of clusters (say K) $\{C_1,C_2,....,C_K\}$ based on some similarity measure. The value of K may or may not be known beforehand. The partition matrix is represented as $U=[u_{kl}]$, $k=1,2,....,K$ and u_{kl}, where u_{kl} is the membership of data point x_j to cluster C_k. For the hard and fuzzy partitioning of the dataset, the conditions that must be satisfied are as follows (Xu & Wunsch, 2009):

$$u_{kl} = \begin{cases} 1 & if\ x_l \in C_k \\ 0 & if\ x_l \notin C_k \end{cases} \tag{1}$$

$$0 < \sum_{l=1}^{n} u_{kl} < n, \quad \forall k \in \{1, 2, ..., K\} \tag{2}$$

Related Works

Though classical clustering algorithms are computationally simple, they have certain shortcomings such as sensitivity towards initialisation of cluster centres and trapping in local optima. There are many algorithms available in the literature to solve these problems such as the evolutional or population-based algorithms. Selim and Al-Sultan (1991) proposed a simulated annealing algorithm for the clustering algorithm. They proved theoretically that a clustering problem's global solution can be reached. Sung and Jin (2000) proposed a tabu search-based heuristic for clustering. They combined the packing and releasing procedures with a tabu search. Krishna and Murty (1999) developed a novel approach called genetic K-Means algorithm for clustering analysis that defines a mutation operator specific to clustering. Maulik and Bandyopadhyay (2000) proposed a genetic algorithm-based method for data clustering problems. The clustering solutions move towards better solutions via selection, crossover and mutation. An ant colony algorithm for clustering was presented by Shelokar et al. (2004). This algorithm mimics the behaviour of ants to find the shortest path from their nest to a food source and back. Its performance was compared with the genetic algorithm, simulated annealing and tabu search and was found to show better performance in comparison to others. Fathian et al. (2007) developed an application of the honey-bee mating optimisation algorithm for data clustering. Particle swarm optimisation (PSO), which simulates the social behaviour of bird flocking, was used for clustering by Kao et al. (2008). Its performance was further enhanced by hybridisation of K-Means with PSO and compared with GA (Murthy & Chowdhury, 1996) and KGA (Bandyopadhyay & Maulik, 2002).

Karaboga and Basturk (2008) described an artificial bee colony (ABC) algorithm for the data clustering problem. Zhang et al. (2010) extended the ABC for data clustering. Its performance was compared with other heuristic-based clustering techniques. Satapathy and Naik (2011) used the teaching learning-based optimisation technique for data clustering. They optimised the cluster centres for a user-specified number of clusters. Hatamlou et al. (2011a) presented the Big Bang-Big Crunch algorithm (BB-BC) for the data clustering problem. In the Big Bang phase, some candidate solutions are randomly generated, which are uniformly spread over the search space. In the Big Crunch phase, some randomly distributed solutions are treated as a single delegate point using a centre of population. Hatamlou et al. (2011b) also presented a gravitational search algorithm-based data clustering technique. Hassanzadeh et al. (2012) presented a firefly algorithm for data clustering. They optimised the cluster centres and extended it to use K-Means clustering to further refine cluster centres. Hatamlou (2013) introduced a new algorithm named Black Hole (BH) algorithm and applied it to solve the clustering problem. Hatamlou and Hatamlou (2013) investigated the combination of GSA and

BB-BC algorithms for clustering problems. They used GSA for exploring the search space for finding the optimal locations of cluster centres and BB-BC was used to diversify the problem. Kumar et al. (2014a) used the gravitational search-based clustering technique for MRI brain image segmentation. Kumar et al. (2014b) also proposed four new variants of harmony search clustering algorithms. They used these variants for solving the clustering problem. They used the search capability of harmony search for optimisation of the within-cluster variation. Saida et al. (2014) presented a cuckoo search for solving data clustering problems. Kumar et al. (2014c) developed a variance-based harmony search algorithm, which is used to solve cluster centre computation problems.

In this paper, differential search-based clustering technique were developed. There are two main reasons for adopting the differential search as a metaheuristic technique for clustering. First is its simplicity, robustness and ease of implementation. Second, it has few control parameters to fine-tune. The proposed approach is applied on real-life datasets.

PROBLEMS AND SOLUTION OF DATA CLUSTERING

Problems and challenges of data clustering

A large number of partitional clustering techniques already exist and are reported in the literature. Some of the well-known clustering techniques are K-Means, Fuzzy C-Means, K-Medoid etc. Among these techniques, the K-Means algorithm is the most widely used clustering technique. It is used for high dimensional datasets due to its simplicity of understanding, ease of implementation and speed of convergence (Duda et al., 2001). Most of the clustering algorithms are sensitive towards the initialisation of cluster centres. Another problem is convergence of the final cluster towards the nearest local optimum solution (Jain, 2010). The K-Means algorithm also suffers from these problems.

Solution to the Problems

Recently, researchers used metaheuristic techniques to overcome problems mentioned in the preceding section. Metaheuristic algorithms are believed to be able to solve NP-hard problems with satisfactory near-optimal solutions with less computational time compared to other classical methods. Although many meta-heuristic algorithms for solving clustering problems have been proposed, the results obtained in terms of accuracy are not up to the mark as reported by Omran et al. (2006), thereby giving rise to the proposal of a DS-based clustering approach.

The proposed approach uses the basic steps of DS. It utilises the concept of population to explore the search space of the given dataset and ensures a greater probability in achieve the near-optimal cluster centres. The cluster centres obtained from the DS algorithm are computed using the inter-cluster distance measure.

Clustering using Differential Search Algorithm

The DS algorithm is used for evolving a set of candidate cluster centres for a fixed number of clusters and determining a near optimal partitioning of the dataset. It is also able to cope with local optima by comparing several solutions simultaneously. The well-known clustering criterion is the sum of squared Euclidean distances between data points, taken individuallys, and the centre of the cluster belonging to every such allocated data point. It is described mathematically as (Xu & Wunsch, 2009):

$$\sum_{i=1}^{n}\sum_{j=1}^{K}\left\| x_i - m_{ij} \right\|^2 \tag{3}$$

where m_{ij} refers to the cluster centre vector of the i^{th} member in j^{th} cluster (say C_{ij}).

The main goal of the proposed DSA-based clustering technique is to search for the appropriate cluster centres such that the above-mentioned clustering criterion is minimised.

Differential Search Algorithm

Civicioglu (2012) developed a novel metaheuristic technique, namely, the differential search algorithm (DS). It is inspired from the migration of living beings, which comprise super-organisms during the climate change of the year. In the DS algorithm, the search space is simulated as the food areas and each point in the search space corresponds to an artificial-super-organism migration (Liu, 2014). The goal of the migration process is to find a global optimal solution to the problem. During this process, the artificial-super-organism chooses the randomly selected positions to be retained temporarily. The members of artificial-super-organisms continue to settle at a position so long as the randomly selected position is suitable for them. As soon as the position becomes unsuitable, members start migrating from that position to another suitable position. The main steps of DS are shown in Figure 1 and described below (Liu, 2014):

In DS algorithm, artificial-super-organisms are made up of artificial-organisms $\left(i.e., X_i, i = \{1, 2, ..., N\}\right)$.

A member of an artificial-organism $\left(i.e., x_{i,j}, j = \{1, 2, ..., D\}\right)$ is randomly initialised using the following equation:

$$x_{i,j} = x_j^{min} + rand \times \left(x_j^{max} - x_j^{min} \right) \tag{4}$$

Here, N indicates the number of artificial organisms in a super-organism and D indicates the size of the respective problem. x_j^{min} and x_j^{max} are the minimum and maximum value of the j^{th} component, respectively.

After initialisation, the mechanism of finding a stopover site $\left(S_i\right)$ at the areas between the artificial-organisms may be described by a Brownian-like random walk model. The algorithm creates a stopover site corresponding to each population individual vector in the current population. The method for producing the stopover site can be described as:

```
Differential Search Algorithm
begin
    Generate initial population of n superorganism, X_i (i = 1,2,....,n)
    Initialize the parameter P_1 and P_2
    Evaluate the fitness of each individual in population
    while (t < MaxGeneration) or (stop criterion)
    Scale = randg(2 × rand) × (rand − rand)
    for i = 1 to n do
        Select randomly select an individual (x_a)
    end for  S_i = x_i + Scale × (x_a − x_i)
    r = rand(N, D)
    if rand < rand then
        if rand < P1 then
            for i = 1 to n do
            r(i,:) = r(i,:) < rand
            end for
        else
            for i = 1 to n do
            r(i, randi(D)) = 0
            end for
        end if
    else
    for i = 1 to n do
    d = randi(D, 1, ⌈P2.rand⌉)
        for j = 1 to size (d, 2) do
        r(i, d(j)) = 0
        end for
    end for
    end if
    r = r > 0
    S(r) = X(r)
    for j = 1 to n do
        Evaluate the offspring S_i
        if S_i is better than X_i then
        X_i = S_i
        end if
    end for
    Store the best solution achieved so far
  end while
end
```

Figure 1. Pseudo code of differential search algorithm (DS)

$$S_i = x_i + Scale \times \left(x_a - x_i \right) \tag{5}$$

where $a \in [1, ..., N]$ is a randomly chosen integer and $a \neq i$. The variable, *scale*, is used to control the size of change in the positions of the individuals of the artificial-organisms. The search process of the stopover site is computed as:

$$S'_{i,j} = \begin{cases} S_{i,j} & if\ r_{i,j} = 0 \\ x_{i,j} & if\ r_{i,j} = 1 \end{cases} \tag{6}$$

where $r_{i,j}$ is an integer number. $S'_{i,j}$ indicates the trail vector. The selection operation helps in selecting the better of the two populations, the stopover site and the artificial-organism as the next population.

DS-Based Clustering Technique

The basic steps of DS, as shown in Figure 1, are used in the proposed clustering technique. In the context of clustering, a member of the artificial super-organism represents the K cluster centres. That is, each member is represented as:

$$x_i = \left[m_{i1}, m_{i2}, ..., m_{ij}, ..., m_{iK} \right] \tag{7}$$

Therefore, a member of the super-organism represents a number of candidate cluster centres for the given dataset. The fitness of each member is computed using Equation 3. The pseudo code of DS-based clustering algorithm is shown in Figure 2.

DS-based Clustering Algorithm

begin
    Initial each individual with K random cluster centres.
    for *iter = 1 to maximum_iteration* ***do***
        for *all individual i* ***do***
            for *all data point X_P in dataset* ***do***
            Compute Euclidean distance of X_P with all cluster centres
            Assign X_P to the cluster that have nearest cluster centre to X_P
            end for
            Compute the fitness function mentioned in Eq.3
        end for
        Update the cluster centres using DS.
    end for
end

Figure 2. Pseudo code of DS-based clustering algorithm

VALIDATION AND DISCUSSIONS

Datasets used

The proposed clustering approach was applied to five benchmark datasets to evaluate its performance (Blake & Merz, 1998). The datasets are *Iris*, *Wine*, *Haberman* and *Contraceptive Method Choice* (*CMC*). Table 1 describes the main characteristics of the used datasets.

Table 1
Description of Datasets Used

Dataset	Number of Instances	Number of Features	Number of Classes	Type
Iris	150	4	3	Real
Wine	178	13	3	Real
Glass	214	9	6	Real
Haberman	306	3	2	Real
CMC	1473	9	3	Real

Performance Metrics

Three well-known cluster quality measures such as precision, recall and G-Measure were used to evaluate the performance of the clustering algorithms. These were defined mathematically for cluster j with respect to class i as follows (Kowalski, 1997):

$$\mathrm{Precision(i,j)} = \frac{\mathrm{N_{ij}}}{\mathrm{N_j}} \tag{8}$$

$$\mathrm{Recall(i,j)} = \frac{\mathrm{N_{ij}}}{\mathrm{N_i}} \tag{9}$$

$$\mathrm{GM}\left(\mathrm{i,j}\right) = \sqrt{\mathrm{Precision}\left(\mathrm{i,j}\right) \times \mathrm{Recall}\left(\mathrm{i,j}\right)} \tag{10}$$

where N_{ij} is the number of data points of class i in the cluster j. N_j is the number of data points of cluster j. N_i is the number of data points of class i. N_T is the total number of cases. A large value of these measures were required for better clustering. The values reported in the tables are average and standard deviation (mentioned in parenthesis) of solutions over 20 independent runs of algorithms.

Algorithms for Comparison

The performance of the DSA-based clustering (DSC) technique was compared with six well-known clustering algorithms such as K-Means (Jain et al., 2010), genetic algorithm-based clustering (GAC) (Maulik & Bandyopadhyay, 2000), modified harmony search-based clustering (MHSC) (Kumar et al., 2014b), particle swarm-based clustering (PSOC) (Omran et al., 2006), flower pollination algorithm-based clustering (FPAC) (Yang, 2012), and bat algorithm-based clustering (BATC) (Yang, 2010).

The population size and maximum number of iterations for all algorithms were set as 30 and 500, respectively. The parameter settings of the above-mentioned algorithms were the same as reported in their original papers.

EVALUATION AND DISCUSSION

Tables 2 to 5 show the comparison between the proposed DS-based clustering approach and above-mentioned techniques for *Iris*, *Wine*, *Haberman* and *CMC* datasets, respectively. For the *Iris* dataset, DSC performed better than the other techniques. The performance of other metaheuristic-based clustering algorithms were almost similar. For the *Wine* dataset, DSC outperformed other clustering techniques. The MHSC was the second best clustering technique among the compared techniques. For *Haberman* and *CMC* datasets, DSC provided better cluster quality measures than the other competitive techniques.

Table 2
Cluster Quality Matrices for Iris Dataset

	KM	GAC	MHSC	PSOC	FPAC	BATC	DSC
Precision	0.3018	0.3534	0.4586	0.4299	0.4266	0.4396	0.6591
	(0.2584)	(0.3193)	(0.3178)	(0.3482)	(0.1936)	(0.4159)	(0.2859)
Recall	0.3020	0.3733	0.4427	0.4460	0.4385	0.4360	0.6466
	(0.2569)	(0.3011)	(0.3126)	(0.3316)	(0.1870)	(0.4090)	(0.2887)
G Measure	0.1144	0.2046	0.4082	0.4364	0.4319	0.4359	0.6502
	(0.2828)	(0.3735)	(0.3454)	(0.3385)	(0.1893)	(0.4107)	(0.2853)

Table 3
Cluster Quality Matrices for Wine Dataset

	KM	GAC	MHSC	PSOC	FPAC	BATC	DSC
Precision	0.2718	0.2925	0.4343	0.3851	0.4074	0.3872	0.4366
	(0.1365)	(0.2191)	(0.2108)	(0.2002)	(0.1290)	(0.2381)	(0.2573)
Recall	0.2742	0.3105	0.4087	0.3757	0.3900	0.3773	0.4302
	(0.1409)	(0.2323)	(0.1897)	(0.1805)	(0.1045)	(0.2233)	(0.2329)
G Measure	0.0777	0.1082	0.2519	0.3795	0.3629	0.3807	0.4263
	(0.2457)	(0.2445)	(0.2940)	(0.1913)	(0.1286)	(0.2296)	(0.2419)

Table 4
Cluster Quality Matrices for Haberman Dataset

	KM	GAC	MHSC	PSOC	FPAC	BATC	DSC
Precision	0.4972	0.4995	0.4996	0.4981	0.4983	0.4941	0.5149
	(0.0141)	(0.0181)	(0.0157)	(0.0144)	(0.0158)	(0.0106)	(0.0195)
Recall	0.4964	0.4994	0.4994	0.4977	0.4978	0.4925	0.5185
	(0.0179)	(0.0229)	(0.0189)	(0.0184)	(0.0206)	(0.0134)	(0.0247)
G Measure	0.4556	0.4647	0.4602	0.4756	0.4764	0.4703	0.4919
	(0.0161)	(0.0226)	(0.0209)	(0.0105)	(0.0241)	(0.0090)	(0.0209)

Table 5
Cluster Quality Matrices for CMC Dataset

	KM	GAC	MHSC	PSOC	FPAC	BATC	DSC
Precision	0.3432 (0.0417)	0.3378 (0.0599)	0.3615 (0.0185)	0.3502 (0.0589)	0.3090 (0.0806)	0.3166 (0.0863)	0.3651 (0.0247)
Recall	0.3320 (0.0444)	0.3372 (0.0570)	0.3569 (0.0184)	0.3472 (0.0514)	0.3124 (0.0748)	0.3186 (0.0827)	0.3551 (0.0290)
G Measure	0.3139 (0.0569)	0.3196 (0.0588)	0.3459 (0.0223)	0.3440 (0.0535)	0.3061 (0.0759)	0.3134 (0.0842)	0.3549 (0.0256)

FUTURE RESEARCH DIRECTIONS

The field of metaheuristic-based partitional clustering is relatively novel and promising with new ideas and applications. Future research directions are as follows.

- The performance of metaheuristic-based clustering algorithms greatly depends upon the setting of algorithm parameters and an encoding scheme. Hence, theoretical analysis is required before the simulation and its implementation.
- The performance consistency of these metaheuristic algorithms need to be ensured.
- Most of the real-life datasets contain data points that are overlapping in nature. There is a need to further explore the possibility of clustering algorithms that take care of overlapped data points and to group them in the appropriate cluster.

CONCLUSION

A novel differential search algorithm-based clustering technique was developed to solve clustering problems. The DSC was applied for data clustering as the number of clusters was known *a priori*. It was compared with four well-known metaheuristic techniques and tested on four datasets. The results revealed that DSC outperformed the four well-known metaheuristic-based clustering techniques.

REFERENCES

Abraham, A., Das, S., & Roy, S. (2008). Swarm intelligence algorithms for data clustering. In O. Maimon, & L. Rokach (Eds.), *Soft Computing for Knowledge Discovery and Data Mining,* pp. 279-313. Berlin: Springer.

Bandyopadhyay, S., & Maulik, U. (2002). An evolutionary technique based on K-means algorithm for optimal clustering. *Information Science*, *146*(1), 221-237.

Blake, C. L., & Merz, C. J. (1998). UCI repository of machine learning. Retrieved from http:/www.ics.uci.edu/_mlearn/databases/.

Civicioglu, P. (2012). Transforming geocentric Cartesian coordinates to geodetic coordinates by using differential search algorithm. *Computer & Geosciences, 46,* 229-247.

Everitt, B. S. (1993). *Cluster analysis*. New York: Edward Arnold and Halsted Press.

Fathian, M., Amiri, B., & Maroosi, A. (2007). Application of honey bee mating optimization algorithm on clustering. *Applied Mathematics and Computation*, *190*(2), 1502-1513.

Frigui, H., & Krishnapuram, R. (1999). A robust competitive clustering algorithm with applications in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(5), 450-465.

Hassanzadeh, T., & Meybodi, M. R. (2012). A new hybrid approach for data clustering using firefly algorithm and K-means, In CSI *International Symposium on Artificial Intelligence and Signal Processing*, 7-11, Shiraz, Fars, Iran.

Hatamlou A., Abdullah, S., & Hatamlou, M. (2011a). Data clustering using big bang-big crunch algorithm. In P. Pichappan, H. Ahmadi, & E. Ariwa (Eds.), *Innovative Computing Technology*. *Lecture notes in communications in computer and information science* (pp. 383-388). US: Springer.

Hatamlou, A., Abdullah, S., & Nezamabadi-pour, H. (2011b). Application of gravitational search algorithm on data clustering, In J. Yao, S. Ramanna, G. Wang, & Z. Suraj (Eds.), *Rough Set and Knowledge Technology*, Lecture notes in computer science (Vol. 6954, pp. 337-346). US: Springer.

Hatamlou, A., & Hatamlou, M. (2013). Hybridization of the gravitational search algorithm and big bang-big crunch algorithm for data clustering. *Fundamental Informaticae*, *126*(4), 319-333.

Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, *222*(1), 175-184.

Jain, A. K. (2010). Data clustering: 50 years beyond *K*-means. *Pattern Recognition Letters*. *31*(8), 651-666.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering a review. *ACM Computing Surveys, 31*(3), 264-323.

Kao, Y. -T., Zahara, E., & Kao, I. W. (2008). A hybridized approach to data clustering. *Expert Systems with Applications, 34*(3), 1754-1762.

Krishna, K., & Murty, M. N. (1999). Genetic k-means algorithm. *IEEE Transactions on System, Man, and Cybernetics, Part B: Cybernetics, 29*(3), 433-439.

Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony algorithm. *Applied Soft Computing, 8*(1), 687-697.

Kumar, V., Chhabra, J. K., & Kumar, D. (2014a). Automatic MRI brain image segmentation using gravitational search based clustering technique. In Dr. R. Srivastava, Dr. S. K. Singh, and Dr. K. K. Shukla (Eds.), *Recent Advances in Computer Vision and Image Processing: Methodologies and Applications* (pp. 313-326). IGI Global, USA.

Kumar, V., Chhabra, J. K., & Kumar, D. (2014b). Parameter adaptive harmony search algorithm for unimodal and multimodal optimization problems. *Journal of Computational Science, 5*(2), 144-155.

Kumar, V., Chhabra, J. K., & Kumar, D. (2014c). Variance-based harmony search algorithm for unimodal and multimodal optimisation problems with application to clustering. *Cybernetics and Systems: An International Journal*, *45*(6), 486-511.

Leung, Y., Zhang, J., & Xu, Z. (2000). Clustering by space-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1396-1410.

Liu, B. (2014). Composite differential search algorithm. *Journal of Applied Mathematics*. Retrieved from http://www.hindawi.com/journals/jam/2014/294703/

Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm based clustering technique. *Pattern Recognition, 33*(9), 1455-1465.

Murthy, C. A., & Chowdhury, N. (1996). In search of optimal clusters using genetic algorithms. *Pattern Recognition Letters, 17*(8), 825-832.

Nanda, S. J., & Panda, G. (2014). A survey on inspired metaheuristic algorithms for partitional clustering. *Swarm and Evolutionary Computation, 16*(2), 1-18.

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2006). Dynamic clustering using particle swarm optimization with application in image segmentation. *Pattern Analysis and Applications, 8*(4), 332-344.

Saida, I. B., Nadjet, K., & Omar B. (2014). A new algorithm for data clustering based on cuckoo search optimization. In J. -S. Pan, P. Kromer, & V. Snasel (Eds.), *Genetic and Evolutionary Computing*. Lecture notes in advances in intelligent systems and computing, Vol. 238, pp. 55-64. Berlin: Springer.

Satapathy, S. C., & Naik, A. (2011). Data clustering based on teaching learning-based optimization. In B. K. Panigrahi, P. N. Suganthan, S. Das, & S. C. Satapathy (Eds.), *Swarm, Evolutionary, and Memetic Computing*. Lecture notes in Computer Science, Vol. 7077, pp. 148-156. Berlin: Springer.

Selim, S. Z., & Al-Sultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern Recognition, 24*(10), 1003-1008.

Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. *Analytica Chimica Acta, 509*(2), 187-195.

Sung, C. S., & Jin, H. W. (2000). A tabu search-based heuristic for clustering. *Pattern Recognition, 33*(5), 849-858.

Xu, R., & Wunsch II, D. C. (2009). *Clustering*. USA: John Wiley and Sons.

Yang, X. S. (2012). Flower pollination algorithm for global optimization. In *Unconventional computation and natural computation* (pp. 240-249). Springer Berlin Heidelberg.

Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In J. R. Gonzalez et al. (Eds.), *Proceedings of Nature Inspired Cooperative Strategies for Optimization,* Studies in Computational Intelligence, Vol. 284, pp. 65-74. Springer: Berlin.

Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. *Expert Systems and Applications, 37*(7), 4761-4767.