Pertanika J. Sci. & Technol. 27 (2): 815 - 827 (2019)

SCIENCE & TECHNOLOGY

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

Vibrating Particle System Algorithm for Hard Clustering Problems

Ashish Parmar[1], Yugal Kumar[1]*, Pradeep Kumar Singh[1] and Vijendra Singh[2]

[1]Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, 173234 India
[2]Department of Computer Science and Engineering, Northcap University, Gurugram, 122017 India

ABSTRACT

In the field of data analysis, clustering is an unsupervised technique that can be used to find identical sets of data. But, it is tough task to find the optimal centroid for a given dataset, especially in hard clustering problems. Recently, a vibrating particle system (VPS) algorithm was developed for solving the optimization problems. This algorithm is based on the concept of free vibration and forced vibration. This algorithm provides more effective and optimal solutions for constrained optimization problems. In this work, the performance of VPS algorithm is evaluated for solving hard clustering problems. The objective of this algorithm is to compute optimal centroid for hard clustering problems. The efficiency of the proposed algorithm is measured on well known clustering datasets and compared with some popular clustering algorithms. The simulation results demonstrate that the VPS algorithm obtains effective results as compared to other algorithms.

Keywords: Clustering, *K*-means, meta-heuristic algorithm, vibrating particle system

INTRODUCTION

Clustering is an unsupervised method to arrange the data into different clusters using distance measure. The data within a cluster is more similar in nature than the other clusters. This method is also used to understand the organisation of data. In clustering, data is partitioned into several clusters based on similarity factor. The similarity factor is based on distance function. The clustering techniques have wide importance in diverse fields such as data analysis, stock market, pattern identification and machine learning paradigms. In literature, different objective functions are reported to find identical group

ARTICLE INFO

Article history:
Received: 13 April 2018
Accepted: 26 November 2018
Published: 25 April 2019

E-mail addresses:
ashishparmar1094@gmail.com (Ashish Parmar)
yugal.kumar.14@gmail.com (Yugal Kumar)
pradeep_84cs@yahoo.com (Pradeep Kumar Singh)
vsingh.fet@gmail.com (Vijendra Singh)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

of data such as Euclidean distance, Manhattan distance, City block distance and Hamming distance. But, it is seen that most of work is reported with Euclidean distance as an objective function. Moreover, clustering is divided into two categories- partitional clustering and hierarchical clustering. In partitional clustering, a fitness function is considered to partition the data into *k* clusters. While in hierarchical clustering, data are merged and split on the basis of objective function. Further, a tree structure is designed to represent clustering results. It is noticed that clustering methods are further categorized into different sub categories based on the nature of data and representations. Some of these are density based clustering, spectral clustering, graph clustering, model based clustering (Hahsler & Bolanos, 2016; Kannan et al., 2004; Murphy & Murphy, 2017; Schaeffer, 2007). Graph clustering is a method in which vertices are grouped together and having minimum number of edges in between clusters. While, model based clustering based on the finite mixture model. This clustering method is implemented on the resulted variable, not on the related covariates. In density based clustering method, clusters are described as maximum set of density connected points. The spectral clustering methods determine the similarity matrix such that all objects lie in singular vector of matrix. But, it is observed that every clustering method requires some similarity measures to find the closeness between data.

Due to technological evolution and wide application area of clustering, many researchers have developed numerous algorithms for solving clustering problems. Large number of meta-heuristic algorithms, swarm based algorithms, evolutionary algorithms and approximation algorithms have been reported in literature. But, none of these algorithms can give exact solution. Hence, according to No Free Lunch Theorem, there is a scope to develop a new algorithm for solving optimization problems that can provide more accurate results. Recently, VPS algorithm is developed for solving the constrained optimization problems (Kaveh & Ghazaan, 2017a). It is seen that this algorithm gives more efficient and optimized results for solving constrained optimization problems. The aim of this work is to examine the effectiveness and performance of the VPS algorithm for solving hard partitional clustering problems. The VPS algorithm is applied to determine the optimal centroid from a given dataset. The performance of the VPS algorithm is evaluated on five real and two artificial datasets. It is seen that the VPS algorithm provides state of art results in comparison to the same class of algorithms.

Related Work

This section describes recent works reported on the clustering problems. Kumar et al. (2016) proposed magnetic charged system search (MCSS) algorithm to handle problem of optimal cluster centers. The proposed algorithm formulates the behavior of charged particles. In the proposed algorithm, Newton's second law of motion is applied for global search. In this study, two artificial and eight real data sets are considered to compute the performance of the MCSS algorithm. It is reported that the proposed algorithm gives effective and efficient clustering results in comparison to *K*-means, genetic algorithm (GA), particle swarm optimization (PSO), and ant colony optimization (ACO) algorithms. A hybrid algorithm based on cat swarm optimization (CSO) and *K*-harmonic means is developed for solving clustering problem (Kumar & Sahoo, 2015a). The CSO algorithm is integrated with *K*-harmonic means to overcome the drawback of local optima. In this work, seven data sets are considered to evaluate the performance of proposed improved CSO (ICSO) algorithm. The performance of the proposed algorithm is compared with existing algorithms and it is concluded that the proposed algorithm resolves the problem of local optima. Moreover, this algorithm also improves convergence speed of CSO algorithm. To handle partitional clustering problems in effective manner, a MCSS-PSO based clustering algorithm is presented (Kumar & Sahoo, 2015b). The performance of the proposed algorithm is evaluated on two artificial and eight real data sets. From simulation results, it is stated that the proposed algorithm is more effective and efficient for handling partitional data clustering problems. An improved version of CSO algorithm is also developed to resolve local optima problem and also improve convergence speed of CSO (Kumar & Sahoo, 2017a). The Cauchy mutation operator is used to prevent local optima problem. The performance of the improved CSO algorithm is evaluated using two artificial and four real data sets. It is seen that the proposed algorithm is more effective and successfully overcome above mentioned problems. A hybrid clustering algorithm based on Monte Carlo equation and Gaussian probability distribution function is also reported to handle local optima problem (Kumar & Sahoo, 2017b). In this work, few benchmark datasets are considered to compute the performance of the proposed algorithm. It is claimed that the proposed algorithm is more effective and robust in order to tackle clustering problems. Han et al. (2017) introduced an enhanced version of gravitational search algorithm, known as bird flock gravitational search algorithm for data clustering. This algorithm is inspired from collective behaviors of birds. In this work, thirteen data sets are considered to measure the performance of the proposed algorithm. The simulation results are compared with several other popular data clustering algorithms. It is seen that the proposed algorithm is one of effective and efficient algorithm for solving clustering problems. Tsai et al. (2017) proposed a clustering framework for cloud data analytics. In this study, eleven data sets are considered to compute the performance of the proposed algorithm. It is observed that the proposed algorithm is more suitable to determine better clustering results in cloud environment. Ozbakır and Turna (2017) presented two novel meta-heuristic algorithms for clustering problems. These algorithms are weighted superposition attraction algorithm and Ion Motion Optimization algorithm. Prior to apply, these algorithms are integrated with the Deb's rule to overcome infeasible solutions problem. It is seen that the proposed algorithm generates more competitive solutions than other algorithms. Kushwaha et al.

(2017) reported a novel clustering algorithm for partitional data clustering. It is based on the concept of magnetic force. The eleven data sets are considered to evaluate the performance of the proposed algorithm. It is observed that the proposed algorithm generates more accurate and robust results as compared to other algorithms. Boushaki et al. (2018), developed an enhanced version of cuckoo search algorithm, known as quantum chaotic cuckoo search algorithm for clustering. The quantum concept is integrated into cuckoo algorithm to resolve local optima problem. In this work, six real life datasets are considered to compute the performance of the proposed algorithm. It is revealed that the proposed algorithm provides better results as compared to other well-known algorithms in terms of internal and external clustering quality. Kumar and Singh (2017) presented an enhanced CSO (ECSO) algorithm for clustering. Further, a local search technique is also incorporated for enhancing the quality of clusters. In this study, five datasets are considered to evaluate the performance of the proposed algorithm. The simulation results are compared with the other existing clustering algorithms. It is reported that ECSO algorithm provides better and enhanced results.

Vibrating Particle System

VPS algorithm is based on the concept of vibration. The term vibration is described using two types- free vibration and forced vibration. In free vibration, the restoring forces are only responsible to maintain the motion. While in forced vibration, a force is applied at certain intervals of time. The frictional effects can be neglected in vibrating system due to undamped vibration. However, due to frictional forces, these vibrations are damped up to some extent. These frictional forces are incepted due to friction between the rigid bodies, dry friction, fluid friction and inter molecular friction (Kaveh & Ghazaan, 2017a; Kaveh & Ghazaan, 2017b). Kaveh and Ghazaan (2017a) presented VPS algorithm based on the concept of free vibration with viscous damping to find global or near global solutions. It is claimed that the proposed algorithm is more convenient and robust in nature. It is a population based meta-heuristic algorithm. Like other meta-heuristic algorithms, the balance between diversification and intensification in VPS is also maintained using particle current population and historical best position. The optimal solution is represented using particle positions. In this algorithm, three equilibrium positions are mentioned with different weights. Further, these positions are updated in each successive generation due to previous best position of the population, known as historically best (*HB)*, good particle (*GP*) and bad particle (*BP*). To determine the *GP* and *BP* from population, the entire population is sorted according to the objective function value. Further, a threshold limit is defined for *GP* and *BP* and finally, *GP* and *BP* are selected randomly from population pool. The particles are initialized in *d*-dimensional search space using Equation (1).

$$X_i^j = x_{min} + r \times (x_{max}^j - x_{min}^j) \tag{1}$$

where X_i^j is the j^{th} variable of the particle i.e. X_{min} represents lower bound vectors, X_{max} denotes upper bound vectors, r is random function in the range of *[0,1]*. To control the effect of the damping level, a decreasing function is also proposed in VPS model. This function is defined in Equation (2). Figure 1 illustrates the vibration motion of a particle with mass. Further, to control the damping effect in vibration, Equation (2) is adopted. This equation can decreases the damping in vibration.

$$D = \left(\frac{iteration}{iteration_{max}} \right)^{\beta} \tag{2}$$

In Equation (2), D represents the damping, iteration is current iteration number, $iteration_{max}$ is maximum number of iteration used for optimization and β is a constant. Further, it is stated that every particle has three equilibrium positions i.e. *HB*, *GP* and *BP* and these positions are updated in each generation. To select good particle or the bad particle for each generation is to sort current population using objective function. Further, a random selection of the *GP* and *BP* is to be made in first and second half respectively. The population of the VPS algorithm is updated using Equations (3-5).

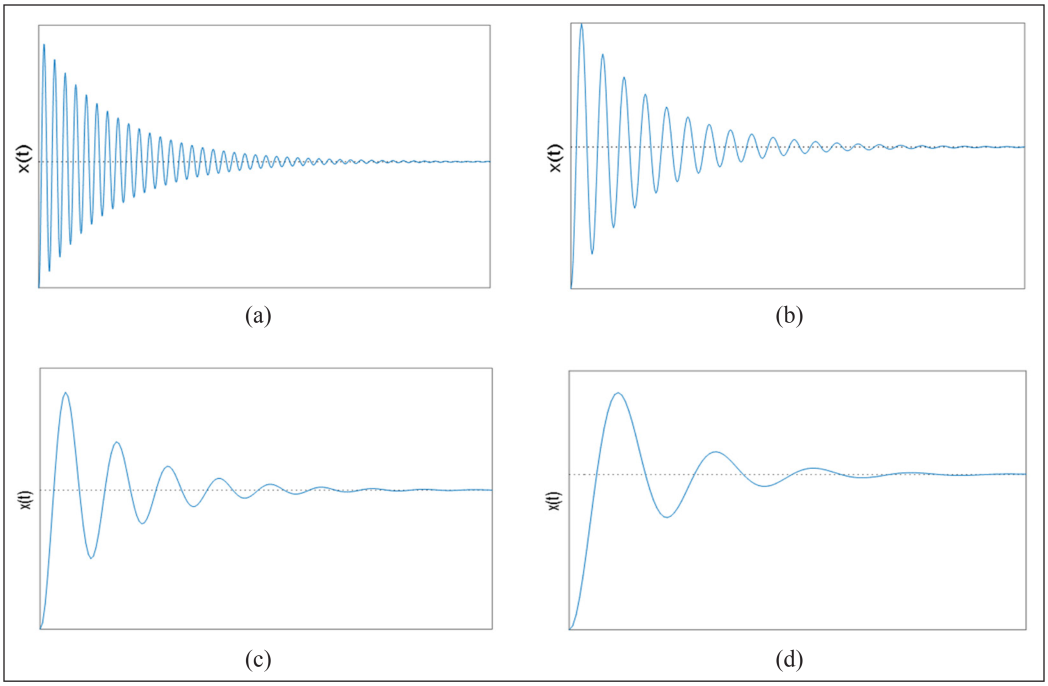

Figure 1. Illustrations of the mechanism of vibration (a-d)

$$x_i^j = w_1 \, [D.A.r_1 + HB^j] + w_2.[D.A.r_2 + GP^j] + w_3.[D.A.r_3 + BP^j] \quad (3)$$

$$A = [w_1.(HB^j - x_i^j] + [w_2.(GP^j - x_i^j)] + [w_3.(BP^j - x_i^j)] \quad (4)$$

$$w_1 + w_2 + w_3 = 1 \quad (5)$$

Here, w_1 is the parameter to measure *HB*, w_2 is the parameter to measure *GP* and w_3 is the parameter to measure *BP*. r_1, r_2 and r_3 are the uniformly distributed random numbers in the range of 0 and 1 and $r_1 \neq r_2 \neq r_3$.

MATERILAS AND METHODS

Proposed VPS Clustering Algorithm

This section describes VPS based clustering algorithm for solving real world clustering problems. The clustering problems are *NP*- hard problem especially when clusters are more than three. The motive of the VPS algorithm is to determine the optimal cluster centroid for hard clustering problems. In clustering problems, the optimal set of clusters is computed using Euclidean distance. In this work, Euclidean distance can be considered as the objective function. It is described using distance between data objects and cluster centres. This function is evaluated for each cluster centres and data objects. Further, the data are associated with the different clusters using minimum Euclidean distance. Euclidean distance is described in Equation (6).

$$minimize\ F(X,C) = \sum_{k=1}^{K} \sum_{x \epsilon D_i} min \parallel X_i - C_k \parallel^2 \quad (6)$$

In Equation (6), X_i denotes the i^{th} data object, C_k represents the k^{th}, and data objects are assigned to clusters according to the minimum distance. A fitness function is also associated with each cluster centres. The fitness function describes the goodness of the clusters. When the data are assigned to clusters, then the value of the function is computed for each cluster centre. In this work, sum of square error (*SSE*) based function is considered to measure the goodness of each cluster centre. This function is defined in Equation (7).

$$F(C_k) = \sum_{k \epsilon 1}^{K} \frac{SSE(C_k)}{\sum_{k=1}^{K} SSE(C_K)} \quad (7)$$

Steps of the VPS Clustering Algorithm. VPS algorithm is a recently developed algorithm, inspired through the behaviour of vibrations. In VPS algorithm, population is represented using particles which are randomly distributed in d-dimensional space. In case of clustering problems, generally the population is described in terms of number of clusters presented in a dataset. The population of VPS algorithm is defined in terms of number of clusters (K) present in a dataset. Further, it is noted that the population should be lie within the boundary constraint (Kaveh & Talatahari, 2010). If boundary condition is violated, then harmony search based approach is adopted to generate the new population within boundary. For clustering problems, the boundary constraints are denoted using minimum and maximum value of each attributes. The main steps of VPS clustering algorithm are listed in Algorithm 1.

Algorithm 1: VPS Clustering Algorithm for Hard Partitional Problems		
Step 1	:	Set up the initial parameters of the VPS algorithm and initialize the initial locations (populations) in an arbitrarily (random) manner
Step 2	:	Compute the objective function values using Equation (6) and also determined the previously best position i.e., *HB* and also compute the fitness function associated with each population.
Step 3	:	To determine the good and bad particles from the population according the fitness function.
Step 4	:	For every particle, compute the values of w_2 and w_3 using random function and satisfied Equation (5).
Step 5	:	Determine the next locations with the help of Equation (3).
Step 6	:	The violated positions are updated using harmony search based mechanism and compute the objective function values for new locations.
Step 7	:	Until the termination condition is reached, repeat steps 3-7
Step 8	:	Obtain final solution generated for hard partitional problems.

RESULTS AND DISCUSION

This section describes the experimental results of VPS clustering algorithm. To investigate the efficiency of the VPS clustering algorithm, some benchmark datasets are considered. These are the well defined datasets and the performance of the newly developed algorithms is tested on these datasets. Further, the effectiveness and efficiency of the algorithm are evaluated using intra cluster distance and standard deviation parameters. The quality of clusters is measured using intra cluster distance parameter and it is measured in terms of best, average and worst. The parameters setting of the proposed VPS clustering algorithm are illustrated in Table 1. The proposed algorithm is implemented in Matlab environment. The experimental results of the VPS algorithm are compared with some other clustering algorithms reported in literature (İnkaya et al., 2015; Kanungo et al., 2016; MacQueen, 1967; Zhan et al., 2009).

Performance Matrices

Intra Cluster Distance. The distance between data objects and their respective cluster centres is known as intra cluster distance. It is used to determine the clustering quality and the results are presented in terms of best, average and worst.

Standard Deviation. Standard deviation is used to determine the information regarding dispersion of data within a cluster. If the value of the standard deviation is low it shows data objects are dispersed near to cluster centres and if the value is high, data is away from centroid.

Table 1
Parameters setting of VPS clustering algorithm

Parameters	Value
Population Size	No. of clusters (K)
β	0.05
w_1	0.3
w_2	0.3
Number of iterations	200

This section describes the results of the proposed algorithm and other meta-heuristic algorithms. Table 2 illustrates the results of the VPS and other clustering algorithms using artificial datasets i.e. ART1 and ART2.

Table 2
Simulation results of the proposed VPS and other clustering results using artificial datasets ART1 and ART2

Dataset	Parameters	*K*-means	PSO	ACO	CSO	TLBO	Proposed VPS
ART 1	Best	157.12	155.46	153.21	153.34	153.96	150.24
	Average	161.12	159.78	157.45	156.54	158.42	156.51
	Worst	166.08	165.34	162.48	160.04	163.07	160.73
	SD	0.846	0.681	0.523	0.689	0.572	0.438
	F-Measure	99.14	100	100	100	100	100
ART2	Best	743	742.26	743.49	738.46	736.12	727.16
	Average	749.83	746.52	747.84	745.17	744.08	741.91
	Worst	754.28	751.03	752.29	750.24	748.59	149.63
	SD	0.726	0.567	0.714	0.498	0.514	0.396
	F-Measure	98.94	99.17	99.08	99.14	99.43	99.48

It is seen that the proposed algorithm gives better results in comparison to other algorithms using all datasets. Further, it is also noticed that the proposed algorithm having minimum intra cluster distance among all other clustering algorithms. The average intra cluster distance of the proposed algorithm is also better than other algorithms. On the analysis on standard deviation parameter, it is observed that VPS algorithm obtains minimum SD values for all datasets.

Table 3
Simulation results of the proposed VPS and other clustering results

Dataset	Parameters	Algorithms					
		K-means	PSO	ACO	CSO	TLBO	Proposed VPS
Iris	Best Case	97.43	96.48	96.89	96.94	96.56	95.63
	Avg. Case	113.08	98.56	98.28	97.86	96.84	95.31
	Worst Case	124.21	99.67	99.34	98.58	98.08	97.79
	SD	16.26	0.467	0.426	0.392	0.546	0.214
	F-Measure	0.782	0.78	0.778	0.781	0.782	0.785
Cancer	Best Case	2991.64	2972.28	2989.12	2978.38	2865.71	2843.41
	Avg. Case	3243.50	3124.09	31848.54	3129.43	3091.44	3045.92
	Worst Case	3614.24	3367.58	3308.17	3456.18	3246.65	3158.64
	SD	256.58	107.14	93.45	128.46	42.11	58.15
	F-Measure	0.832	0.826	0.829	0.831	0.834	0.836
CMC	Best Case	5813.29	5786.81	5746.23	5718.78	5778.61	5648.23
	Avg. Case	5914.46	5837.72	5828.42	5804.52	5836.25	5761.28
	Worst Case	5992.33	5949.47	5941.14	5921.28	5921.32	5873.42
	SD	49.62	48.86	44.34	43.29	38.96	34.56
	F-Measure	0.337	0.333	0.332	0.334	0.331	0.335
Wine	Best Case	16768.18	16483.61	16448.35	16431.76	16578.42	16106.42
	Avg. Case	18061.24	16417.47	16530.53	16395.18	16360.04	16256.82
	Worst Case	18764.49	16594.26	16616.36	16589.54	16917.26	16796. 35
	SD	796.13	88.27	48.86	62.41	56.14	37.83
	F-Measure	0.519	0.516	0.522	0.521	0.52	0.524
Glass	Best Case	222.43	264.56	273.22	256.53	246.89	229.22
	Avg. Case	246.51	278.71	281.46	264.44	256.44	246.38
	Worst Case	258.38	283.52	286.08	282.27	287.52	276.54
	SD	18.32	8.59	6.58	15.43	15.29	5.78
	F-Measure	0.426	0.412	0.402	0.416	0.422	0.426

Table 3 demonstrates the experimental results of VPS and other state of art clustering algorithms using real life datasets. In this study, five datasets are taken into consideration. These datasets are iris, cancer, CMC, wine and glass. It is seen that proposed VPS clustering algorithm obtains minimum intra cluster distance among all other algorithms. On the analysis of F-measure parameter, it is also stated that the proposed algorithm gave better performance as compared to other algorithms. On the behalf of simulation results, it can be concluded that in VPS algorithm, objects within clusters are tightly bound than other algorithms being compared.

Figure 2 shows the distribution of the objects in wine dataset using malic acid and alcohol attributes. While, Figure 3 shows the clustering of objects in wine dataset into different clusters using VPS clustering algorithm. Figure 4 depicts the distribution of wine dataset using malic acid, alcohol, and ash attributes. Figure 5 depicts the clustering results of VPS algorithm. Figure 6 illustrates the convergence behaviour of proposed VPS, ACO, *K*-Means and PSO clustering algorithms using wine dataset. It can be concluded that convergence rate of the proposed algorithm is better than other algorithms. Finally, it is concluded that the proposed algorithm is more capable and efficient than other algorithm being compared.

Figure 2. Distribution of data objects in Wine dataset (2D View)

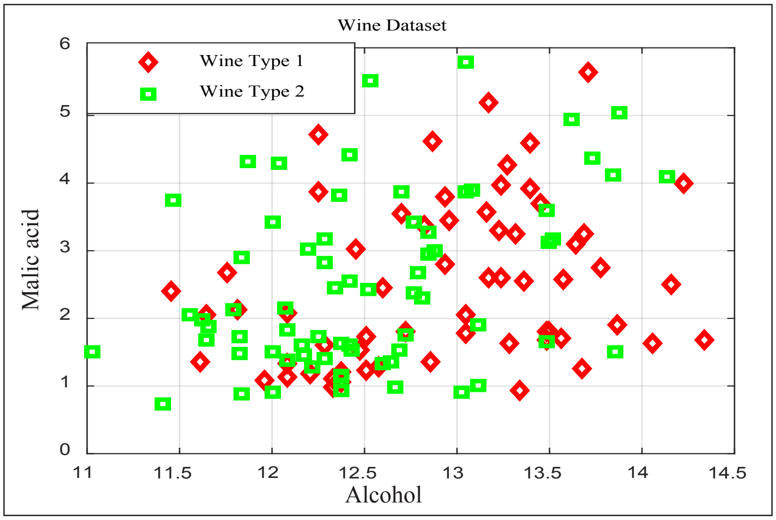

Figure 3. Clustering results of VPS algorithm using Wine dataset (2D View)

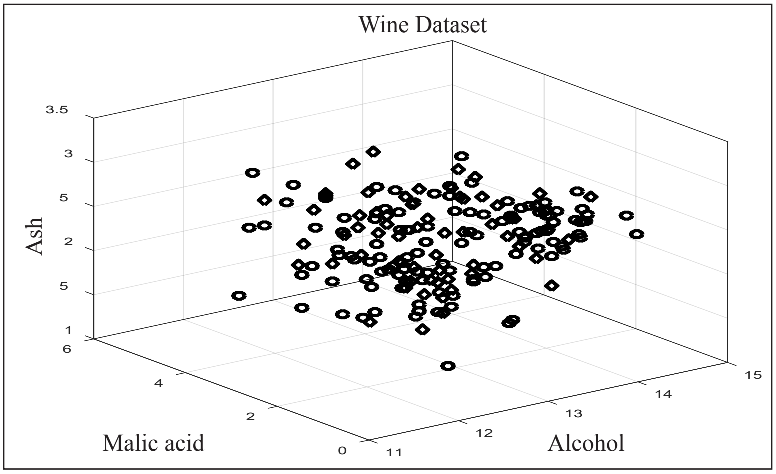

Figure 4. Distribution of data objects in Wine dataset (3D View)

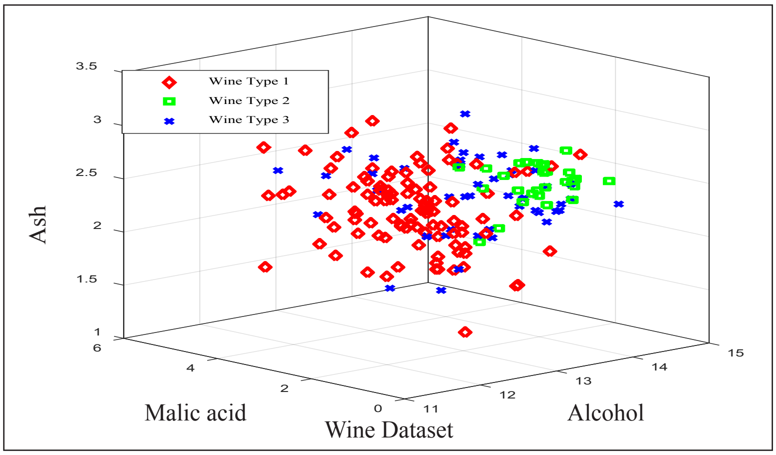

Figure 5. Clustering results of VPS algorithm using Wine dataset (3D View)

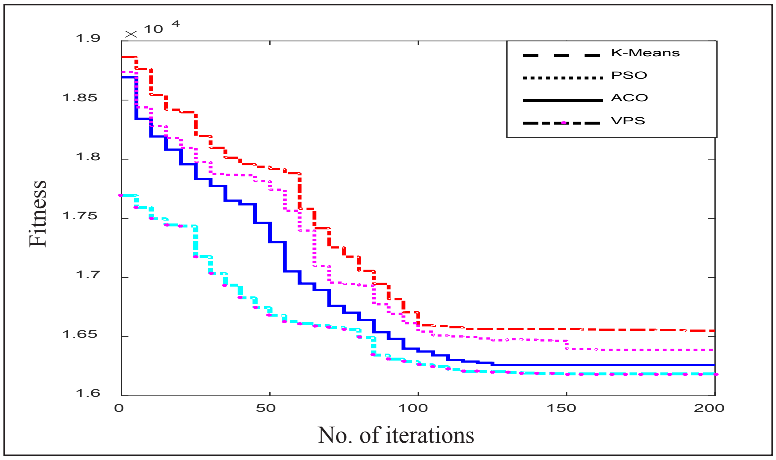

Figure 6. Convergence results of VPS algorithm using Wine dataset

CONCLUSION

In this work, VPS algorithm is proposed for solving hard clustering problem. This algorithm is based on the concept of free vibration. The VPS algorithm is adopted to determine optimal cluster centroid and also minimize intra cluster distance for hard clustering problem. The efficacy of the proposed algorithm is tested on benchmark clustering datasets. Further, intra cluster distance and standard deviation parameters are used as performance parameters. The simulation results showed that VPS algorithm obtains better results than other existing cluttering algorithms. It is concluded that the proposed algorithm is more effective, efficient and robust for solving hard clustering problem.

REFERENCES

Boushaki, S. I., Kamel, N., & Bendjeghaba, O. (2018). A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications, 96*, 358-372.

Hahsler, M., & Bolaños, M. (2016). Clustering data streams based on shared density between micro-clusters. *IEEE Transactions on Knowledge and Data Engineering, 28*(6), 1449-1461.

Han, X., Quan, L., Xiong, X., Almeter, M., Xiang, J., & Lan, Y. (2017). A novel data clustering algorithm based on modified gravitational search algorithm. *Engineering Applications of Artificial Intelligence, 61*, 1-7.

İnkaya, T., Kayalıgil, S., & Özdemirel, N. E. (2015). Ant colony optimization based clustering methodology. *Applied Soft Computing, 28*, 301-311.

Kannan, R., Vempala, S., &Vetta, A. (2004). On clusterings: Good, bad and spectral. *Journal of the ACM (JACM), 51*(3), 497-515.

Kanungo, D. P., Nayak, J., Naik, B., & Behera, H. S. (2016). Hybrid clustering using elitist teaching learning-based optimization: an improved hybrid approach of TLBO. *International Journal of Rough Sets and Data Analysis (IJRSDA), 3*(1), 1-19.

Kaveh, A., & Ghazaan, M. I. (2017a). A new meta-heuristic algorithm: Vibrating particles system. Scientia Iranica. *Transaction A, Civil Engineering, 24*(2), 551-566.

Kaveh, A., & Ghazaan, M. I. (2017b). Vibrating particles system algorithm for truss optimization with multiple natural frequency constraints. *Acta Mechanica, 228*(1), 307-322.

Kaveh, A., & Talatahari, S. (2010). A novel heuristic optimization method: charged system search. *Acta Mechanica, 213*(3-4), 267-289.

Kumar, Y., & Sahoo, G. (2015a). A hybrid data clustering approach based on improved cat swarm optimization and K-harmonic mean algorithm. *Ai Communications, 28*(4), 751-764.

Kumar, Y., & Sahoo, G. (2015b). Hybridization of magnetic charge system search and particle swarm optimization for efficient data clustering using neighborhood search strategy. *Soft Computing, 19*(12), 3621-3645.

Kumar, Y., & Sahoo, G. (2017a). An Improved Cat Swarm Optimization Algorithm Based on Opposition-Based Learning and Cauchy Operator for Clustering. *JIPS (Journal of Information Processing Systems), 13*(4), 1000-1013.

Kumar, Y., & Sahoo, G. (2017b). Gaussian cat swarm optimisation algorithm based on Monte Carlo method for data clustering. *International Journal of Computational Science and Engineering, 14*(2), 198-210.

Kumar, Y., & Singh, P. K. (2017). Improved cat swarm optimization algorithm for solving global optimization problems and its application to clustering. *Applied Intelligence, 48*(9), 2681-2697. doi: 10.1007/s10489-017-1096-8.

Kumar, Y., Gupta, S., Kumar, D., & Sahoo, G. (2016). A clustering approach based on charged particles. In O. Baskan (Ed.), *Optimization Algorithms-Methods and Applications* (pp. 245-263). London: IntechOpen.

Kushwaha, N., Pant, M., Kant, S., & Jain, V. K. (2017). Magnetic optimization algorithm for data clustering. *Pattern Recognition Letters, 115*, 59-65.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297). London: Cambridge University Press.

Murphy, K., & Murphy, T. B. (2017). Parsimonious Model-Based Clustering with Covariates. In *Statistics archive* (pp. 1-17). Ithaca: Cornell University. doi: arXiv: 1711.05632.

Özbakır, L., & Turna, F. (2017). Clustering performance comparison of new generation meta-heuristic algorithms. *Knowledge-Based Systems, 130*, 1-16.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review, 1*(1), 27-64.

Tsai, C. W., Liu, S. J., & Wang, Y. C. (2017). A parallel metaheuristic data clustering framework for cloud. *Journal of Parallel and Distributed Computing, 116*, 39-49.

Zhan, Z. H., Zhang, J., Li, Y., & Chung, H. S. H. (2009). Adaptive particle swarm optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39*(6), 1362-1381.