Modified adaptive inertia weight particle swarm optimisation for data clustering

Vikash Yadav* and Indresh Kumar Gupta

Department of Computer Science and Engineering,
ABES Engineering College, Ghaziabad,
Uttar Pradesh, India
and
Harcourt Butler Technical University,
Nawabganj, Kanpur,
Uttar Pradesh, India
Email: vikas.yadav.cs@gmail.com
Email: indresh.gupta345@gmail.com
*Corresponding author

Abstract: Data clustering is widely applied in many real world domain including marketing, anthropology, medical science, engineering, economics, and others. It concerns with the partition of unlabelled dataset objects into clusters (groups) based on a similarity measure. The partitioning approach of dataset objects must follow that intra-cluster distances are smaller and inter-cluster distances are larger. In the current work, particle swarm optimisation (PSO) is employed for clustering. Some times the PSO may get stuck into a local optima; to overcome the PSO algorithm's trapping in a local optima a modified adaptive inertia weight particle swarm optimisation (MAIWPSO) is developed for data clustering based on fitness value of particles. K-means, PSO and MAIWPSO for clustering have been simulated on six standard dataset namely iris, thyroid, heart, breast cancer, crude oil and pima. Simulation results confirm MAIWPSO is a better approach for clustering against K-means and PSO.

Keywords: data clustering; K-means clustering; particle swarms optimisation; PSO; fitness function.

Reference to this paper should be made as follows: Yadav, V. and Gupta, I.K. (2022) 'Modified adaptive inertia weight particle swarm optimisation for data clustering', *Int. J. Innovative Computing and Applications*, Vol. 13, No. 1, pp.34–40.

Biographical notes: Vikash Yadav received his BTech in Computer Science and Engineering in 2009 from Uttar Pradesh Technical University, Lucknow, U.P., India, MTech in Software Engineering in 2013 from Motilal Nehru National Institute of Technology, Allahabad, U.P., India and PhD in Computer Science and Engineering from Dr. A.P.J Abdul Kalam University (Formerly U.P. Technical University) Lucknow, U.P., India in 2017, in the field of image processing. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, India and has more than eight years of teaching/research experience and published more than 30 research papers in various national/international conferences/journals. His areas of interest include data structure, data mining, image processing and big data analytics.

Indresh Kumar Gupta is currently working as a Guest Lecturer in the Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, India. He completed his MTech degree in Computer Science and Engineering from ABV IIITM Gwalior and BTech degree from AKTU Lucknow. His areas of research include nature-inspired optimisation techniques, machine learning, data mining and information security. He has published several papers in various international conferences/journals.

1 Introduction

Clustering is a crucial approach to discover inherent structure in a dataset by grouping the objects of dataset into subsets which have some meaningful information in context of particular problem. Clustering is used in various field ranging from life science, social science, computer science and engineering etc. (Fahad et al., 2014; Lin et al., 2005; Otto et al., 2018). It is an act of partitioning an unlabelled dataset objects into number of groups such that objects within the same group have high similarity than to those of other groups. Each group is called cluster. A number of similarity measures are proposed in past few decades (Jaskowiak et al., 2013). In this paper we use euclidean distance as similarity measure. The euclidean distance for any two objects p and q is given as:

$$Dist\left(O_p, O_q\right) = \sqrt{\sum_{l=1}^{D}\left(O_{pl} - O_{ql}\right)^2} \tag{1}$$

In equation (1) O_{pl} and O_{ql} denote the value of l^{th} feature for p^{th} and q^{th} object respectively and D stands for total features of object.

On the basis of similarity measure clustering technique partitioned the objects of dataset $O = \{O_1, O_2, O_3 \ldots . O_n\}$ into K clusters $C = (C_1, C_2, C_3 \ldots . C_K)$. Thus,

$$\bigcup_{i=2}^{K} C_i = O \tag{2}$$

$$C_i \cap C_j = \phi \quad for\ i, j \in \{1, 2, 3 \ldots . K\} \tag{3}$$

$$C_i \neq \phi \qquad for\ i \in \{1, 2, 3 \ldots \ldots K\} \tag{4}$$

The above equations (2) to (4) clearly depict each data object must fall at least one cluster and none of the cluster to be empty.

In current paper a modified adaptive inertia weight particle swarm optimisation (MAIWPSO) is developed for data clustering to enhance the performance. The primary aim of clustering technique to discover optimal centre for K clusters in such a way that sum of intra cluster distance is minimum. Thus clustering metric and objective function is defined as follows:

$$Clustering\ metric(\psi) = \sum_{i=1}^{K} \psi_i \tag{5}$$

$$\psi_i = \sum_{O_m \in C_i} W_{mi} Distance\left(O_m, C_i\right) \tag{6}$$

$$Objective\ function\,(obj) = \min(\psi) \tag{7}$$

In equation (6) *Distance* (O_m, C_i) stands for the similarity between i^{th} cluster centre and object m; W_{mi} is set to 1 if object m falls to i^{th} cluster otherwise it is set to 0.

2 Related literature

Chen and Ye (2012) developed a clustering technique based on original PSO called PSO-clustering. They employed the euclidean distance to compute the similarity between centre of cluster and object. Traditional approach like K-means suffers from the optimal solution due to the number of clusters to be known at begging. Thus these algorithms can't ensure to give optimal solution each time. PSO take few parameters as compared to other evolutionary algorithm. To verify the performance of K-means, fuzzy C-means (FCM) and PSO-clustering, they implement all approaches on four artificial datasets. Simulation results confirm PSO-clustering has better performance against traditional clustering techniques.

Omran et al. (2006) proposed a dynamic clustering technique (DCPSO) based on PSO. DCPSO itself discover optimimum number of clusters and simultaneously generate the clusters with minimum user interaction. The key feature is user don't worry about the how many number of clusters to be formed, since knowing the number of clusters to be formed at beginning is very difficult task and affect the quality of clusters. This algorithm integrates PSO and K-means. DCPSO start with partitioning objects into more clusters, and then Binary PSO is employed to discover best number of clusters. At next step K-means algorithm is employed to refine the centre of selected clusters. A validity index measure is used as performance measure for clusters obtained. This novel algorithm was applied on natural and synthetic images. Simulation confirm algorithm discover optimum number of clusters and a better approach for clustering.

Ghali et al. (2009) developed a clustering approach based on exponential particle swarm optimisation (EPSO). EPSO employed exponential inertia weight in lieu of linear inertia weight. They compared both techniques of clustering on five standard dataset namely breast cancer, iris, yeast, glass and lenses. Simulation results confirm clustering based on EPSO has lower quantisation error than PSO but EPSO slowly converge in comparison of PSO. EPSO algorithm for clustering increased the probability to discover suitable clusters centre as it decrease the number of failure.

Esmin and Matwin (2012) proposed a clustering approach based on particle swarm optimisation (PSO) with mutation operator. This hybrid particle swarm optimisation with mutation (HPSOM) employs PSO and integrates mutation operation of GA often used to tackle the problem of local optima. Authors implement K-mean, PSO and HSPOM for clustering problem on five dataset namely artificial problem, iris, wine and breast cancer. For fare comparison of all algorithm each algorithm runs 30 times. Comparison of HPSOM with PSO and K-means showed that HPSOM has least fitness error with better convergence. New HPSOM has smaller intra-cluster and larger inter cluster distances. Simulation results confirm HPSOM algorithm for clustering is efficient and generate compact clusters.

Izakian et al. (2009) developed a clustering technique based on FCM and fuzzy particle swarm optimisation (FPSO). FCM algorithm is straight forward, easy to implement and efficient, but it may stuck in local optima because it is sensitive to the initialisation, while PSO is global stochastic tool which is applied to large number of optimisation problem. Author's combines the FCM and FPSO to take the advantage of both techniques. FCM, FPSO and FCM-FPSO were implemented on popular dataset namely iris, wine, contraceptive method choice (CMC) and vowel. Experimental results prove that FCM-FPSO produced high quality clusters.

Tsai and Kao (2011) developed a novel PSO with selective particle regeneration (SRPSO). This algorithm has two main characteristics:

- Unbalanced parameter setting and particle regeneration operation. Key role of unbalanced parameter setting to provide faster convergence to the algorithm while particle regeneration operation employed to suit out the problem to stuck into local optimal point and generate particle with better fitness value. Authors developed SRPSO and KSRPSO (K-means PSO) techniques for clustering. They simulated PSO, SRPSO and KSRPSO on two artificial and seven standard dataset namely, crude oil, iris, vowel, cancer, wine, glass and CMC. Simulation result confirms proposed clustering techniques have better performance against K-means, original PSO and traditional clustering methods. Authors also examined the proposed clustering techniques against other improved KNM-PSO (Kao et al., 2008), KGA (Bandyopadhyay and Maulik, 2002), PSO+R1 (Kao et al., 2007), and PSO+R2 (Kao et al., 2007) with respect to intra-cluster distances and standard deviation and computational result confirm both approaches are better in terms of these two measures.

Li et al. (2015) proposed a novel approach for clustering based dynamic PSO with K-means (DPSOK) for image segmentation. K-means algorithm for clustering suffers from optimum solution as it highly dependent on initial number of clusters to be formed. PSO is widely used as heuristic optimisation approach and sometime it may stuck in local optima. In PSO inertia weight is an important factor, Authors incorporate dynamic inertia weight and learning factor to maintain the exploration and exploitation trade off, i.e., to maintain equilibrium optimisation capability. PSOK, K-means and DPSOK were implement on five popular image dataset namely panda, smile, bridge, waterfall, cymbidim and tea. Experimental results depict that DPSOK outperformed the classic K-means and PSOK algorithm.

Sengupta et al. (2018) proposed a clustering approach based on hybridisation of quantum behaved PSO and fuzzy c-means (QPSO FCM). Novel PSO employ fully connected topology. The main use of QPSO to escape from stagnation in local optima while FCM is use to partition data objects depending upon membership probabilities. Authors simulate QPSO FCM, PSO K-means and QPSO K-means on five dataset namely breast cancer, iris, seeds, sonar and mammographic mass. They use criterion namely F-measure, accuracy, intra cluster distance, inter cluster distance and quantisation error for estimation of performance. Simulation results indicate QPSO FCM for clustering produced better results compare to both approaches.

Bouyer and Hatamlou (2018) developed a clustering approach based on K-harmonic means (KHM), improved cuckoo search (ICS) and PSO. This novel algorithm for clustering is called ICMPKHM. The main use of KHM to overcome the problem of sensitivity of initialisation of K means. ICS produce optimal result on employing levy flight distribution. Thus this novel ICS has high convergence than original cuckoo search. ICS is combined with PSO to overcome the problem of local optima. Author simulate original PSO, KHM, PSOKHM, and ICAKHM on two artificial dataset and eight real dataset namely iris, wine, wisconsin breast cancer, glass, CMC, thyroid gland, vowel, and ecoli. They use criterion namely mean square error, F measure, average standard deviation, intra cluster distance, Davies-Bouldin (DB) index and silhouette coefficient for estimation of performance. Simulation results indicate that ICMPKHM is not sensitive to initialisation of centroid and produce high quality clustering results compared to other algorithms. Some other good research on clustering based on PSO is given in papers (Alswaitti et al., 2018; Saha and Das, 2018; Izakian et al., 2016).

3 PSO for clustering

PSO is a computational optimisation technique developed by Eberhart and Kennedy (1995) and Zhan et al. (2009). It is population based evolutionary algorithm which can mimic the fish schooling and bird flocking behaviour. In this technique each solution is called particle and population is called swarm (Gupta et al., 2019; Parveen et al., 2018). During the searching each particle fly in multidimensional solution space of problem and change their position based on own experience and influence of swarm. Every particle has limited memory, which is used to keep record of current velocity, current position, fitness value (quality), own best position and swarm best position. We denote the current velocity of i^{th} particle as $V_i^t = [v_{i1}^t, v_{i2}^t, v_{i3}^t.....v_{iD}^t]$ and current position as $X_i^t = [x_{i1}^t, x_{i2}^t, x_{i3}^t.....x_{iD}^t]$. In PSO i^{th} particle best position is denoted as $P_i^t = [p_{i1}^t, p_{i2}^t, p_{i3}^t.....p_{iD}^t]$ and swarm best position is denoted as $G^t = [g_1^t, g_2^t, g_3^t.....x_{iD}^t]$. The velocity and position of i^{th} particle are updates using following equations:

$$v_{id}^{t+1} = w.v_{id}^t + c_1 r_1^t \left[p_{id}^t - x_{id}^t \right] + c_2 r_2^t \left[g_d^t - x_{id}^t \right] \quad (8)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (9)$$

In equation (8), $d = \{1, 2, 3\ldots D\}$ denote the dimension of search space. Role of cognitive acceleration constant c_1 to magnetise the particle movement towards own best position while role of social acceleration constant c_2 to magnetise the movement of particle towards swarm best position. $r_1, r_2 \in [0, 1]$ are two uniformly distributed number. w stands for inertia weight which was given by Shi and Eberhart (1998). They linearly decrease the value of inertia weight according to below equation:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{t_{\max}} * t \quad (10)$$

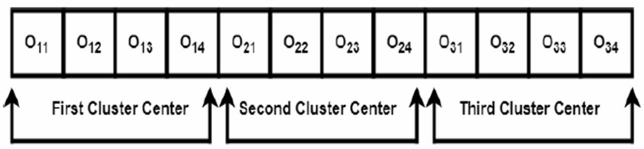
In equation (10) t_{max} is predefined number of iteration and t represent current iteration. Shi and Eberhart set the value of $W_{max} = 0.9$ and $W_{min} = 0.4$. This inertia weight approach for PSO facilitates exploration at initial iterations and exploitation at later iterations. Large value of inertia weight promotes exploration while its small value promotes exploitation.

However if particle fall into local optimal position then there may be the chance it can't dispose itself from that position. For instance if best position obtained by swarm is local optimal, personal and current best position of i^{th} particle is also local optimal then the second and third term of equation (8) incline towards zero. Inertia weight w is linearly decreasing, thus velocity of i^{th} particle at next iteration is approximate to zero, and particle's position at next iteration is not updated. Thus particle still in local optimal position. To overcome the particle's trapping in local optimal position a modified adaptive inertia weight is developed for PSO based on fitness value of particles and applied to the clustering.

3.1 *Particle representation*

For particle representation real encoding is used. Each particle is a sequence of real valued number denoting the K clusters centre. For dataset with D dimensional features, the length of each particle is $K \, X \, D$, where first D numbers of sequence denote the features of first cluster centre. Next D numbers of sequence denote the features of second cluster centre and so on. A particle with 4 features and 3 clusters is shown in Figure 1.

Figure 1 Particle representation in PSO for clustring



3.2 *Fitness evaluation*

Fitness Computation is carried out in two steps. In first step based on the cluster centres each object O_m, $m = 1, 2\ldots\ldots N$ is assign to the i^{th} cluster having centre Ci as:

$$Distance(O_m, C_i) \le Distance(O_m, C_j) \quad j = 1, 2, \ldots K \text{ and } j \ne i \tag{11}$$

New centre for i^{th} cluster is computed as:

$$C_i^* = \frac{\sum_{m=1}^{N} W_{mi} O_m}{\sum_{m=1}^{N} W_{mi}} \quad i = 1, 2, 3 \ldots K \tag{12}$$

After obtaining new clusters centre we can compute the intra cluster distance ψ_i for i^{th} cluster using equation (6). Fitness function is reciprocal of intra cluster distance. Thus fitness value of i^{th} particle is given as follows:

$$fit_i = \frac{1}{\psi_i} \tag{13}$$

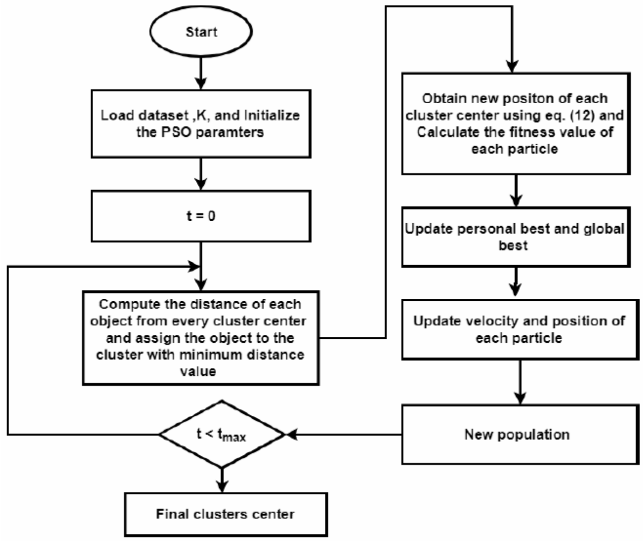
3.3 *Selection of best positions*

Best positions of particles at each iteration highly affect the quality of solution. Thus proper selection of these two positions is essential. If at current iteration fitness value of particle's personal best position is less than the fitness value of particle's current position, then personal best position is set to current position otherwise there is no change. Highest fittest personal best position of any particle is set as swarm best position.

3.4 *Updation of particle velocity and position*

These are two final steps of PSO. We update the i^{th} particle velocity using equation (8) and position using equation (9).

Figure 2 Flow chart of PSO for clustering



4 MAIWPSO for clustering

In current paper a modified inertia weight is developed for clustering. The new PSO use particle fitness value in computation of inertia weight. Suppose G_t denote the global best position of swarm at iteration t and its corresponding fitness value is fit_G^t. Thus global best solution fitness differential between t^{th} and $t-1^{th}$ can be given as:

$$\Delta fit^t = \left| fit_G^t - fit_G^{t-1} \right| \tag{14}$$

Thus convergence speed is defined as:

$$\tau = \frac{\Delta fit^t}{fit_{\max}} \tag{15}$$

In equation (15) $fit_{\max} = \max \{\Delta fit^1, fit^2, \ldots fit^t\}$. Therefore convergence speed τ is greater than 0 and less than or equal to 1.

In PSO it is very essential to maintain the diversity of swarm during the course of searching. The derivation philosophy can be applied to preserve the diversity of swarm. This approach for PSO can be defined as:

$$fit_{avg}^{t} = \frac{1}{Popsize} \sum_{i=1}^{Popsize} fit_i^{y} \quad (16)$$

$$fit_{norm}^{t} = \max \left\{ \begin{matrix} \left| fit_1^t - f_{avg}^t \right|, \left| fit_2^t - f_{avg}^t \right|, \\ \ldots\ldots \left| fit_{Popsize}^t - f_{avg}^t \right| \end{matrix} \right\} \quad (17)$$

$$\sigma^{t} = \frac{1}{Popsize} \sum_{i=1}^{Popsize} \frac{\left| fit_i^t - fit_{avg}^t \right|}{fit_{norm}^t} \quad (18)$$

In equation (16) to equation (18) fit_{avg}^t denote the average fitness value of swarm at iteration *t*. fit_i^t stands for fitness value for i^{th} particle at iteration *t*. fit_{norm}^t denote the normalisation factor of fitness and σ^t represents the swarm diversity. Swarm diversity σ^t is greater than zero and less than or equal to 1.

Modified adaptive inertia weight integrates the convergence speed τ with respect to best fitness value obtained so far at current iteration *t* and swarm diversity σ^t with respect to population deviation. Thus modified adaptive inertia weight is defined as:

$$w = \theta.\tau + \lambda.\sigma^{t|} + w_{\min} \quad (19)$$

In equation (19), θ, λ are factors and τ, σ^{τ} are greater than 0 and less than or equal to 1. So, $w_{\min} \leq w \leq \theta. + \lambda + w_{\min}$. As Shi and Eberhart set the value of $w_{\max} = 0.9$ and $w_{\min} = 0.4$ so in equation (19) we tune $w_{\min} = 0.4$ and $\theta + \lambda + w_{\min} = 0.9$.

5 Computational results and discussion

5.1 Computational environment

PSO and MAIWPSO for clustering is simulated on six standard datasets selected from UCI database repository (https://archive.ics.uci.edu/ml/index.php). The selected dataset includes iris, thyroid, heart, crude oil and pima. Table 1 provides the characteristic of selected dataset including the number of features, number of clusters and total objects. Matlab 7.1 is used as implementation software of this paper. Population size *Pop size* is taken to 50, acceleration constant c_1, c_2 are set to 2. Each experiment used 500 generations.

Table 1 Characteristic of datasets

Sr. no.	*Dataset*	*No. of features*	*No. of clusters*	*Total objects*
1	Iris	4	3	150
2	Thyroid	5	3	215
3	Heart	14	2	303
4	Breast cancer	9	2	683
5	Crude oil	5	3	56
6	Pima	8	2	768

5.2 Computational results and analysis

We select clustering metric (sum of intra cluster distance), error rate and computational time as criterion for estimation of clustering algorithms. Percentage of misplaced data objects is defined as error rate. Computational outcomes are given in Tables 2 to 4. There are 10 columns in each table. Column 2 stands for serial number, column 2 denote problem instance, columns 3 to 7 has clustering metric and next column 8 denote the mean of clustering metric. Column 9 stands for mean error rate and mean computational time. We noted mean of clustering metric (sum of intra cluster distance) is least for every dataset in case of MAIWPSO compared to PSO and K-means. Average means error rate for MAIWPSO, PSO and K-mean is 11:67%, 15:05% and 17:09% while average computational time for MAIWPSO, PSO and K-means is 26.07 seconds, 20.14 seconds and 62.09 seconds. Average mean error rate for MAIWPSO is very less in comparison of PSO and K-means while MAIWPSO takes more computational time in comparison of PSO. Graphical comparison of mean error rate and mean computational time w.r.t problem instance is shown in Figures 3(a) and 3(b) respectively.

Table 2 Result of K-means for clustering problems

Sr. no.	*Test problem*	*Number of runs*					*Mean*	*% Error*	*Time (sec)*
		1	*2*	*3*	*4*	*5*			
1	Iris	116.73	109.28	103.73	100.83	118.03	109.72	20.49	25.40
2	Thyroid	1,998.82	2,000.68	1,976.23	1,966.39	1,959.83	1,980.39	16.64	36.22
3	Heart	1,698.79	1,704.92	1,700.33	1,762.28	1,716.33	1,716.53	18.13	52.28
4	Breast cancer	3,281.82	3,314.12	3,301.41	3,352.80	3,340.32	3,318.09	9.05	116.68
5	Crude oil	278. 68	279.88	279.89	278.96	281.49	279.78	18.48	9.41
6	Pima	2973.42	2,898.72	2,886.05	2,942.41	2,890.05	2,918.13	19.77	132.57

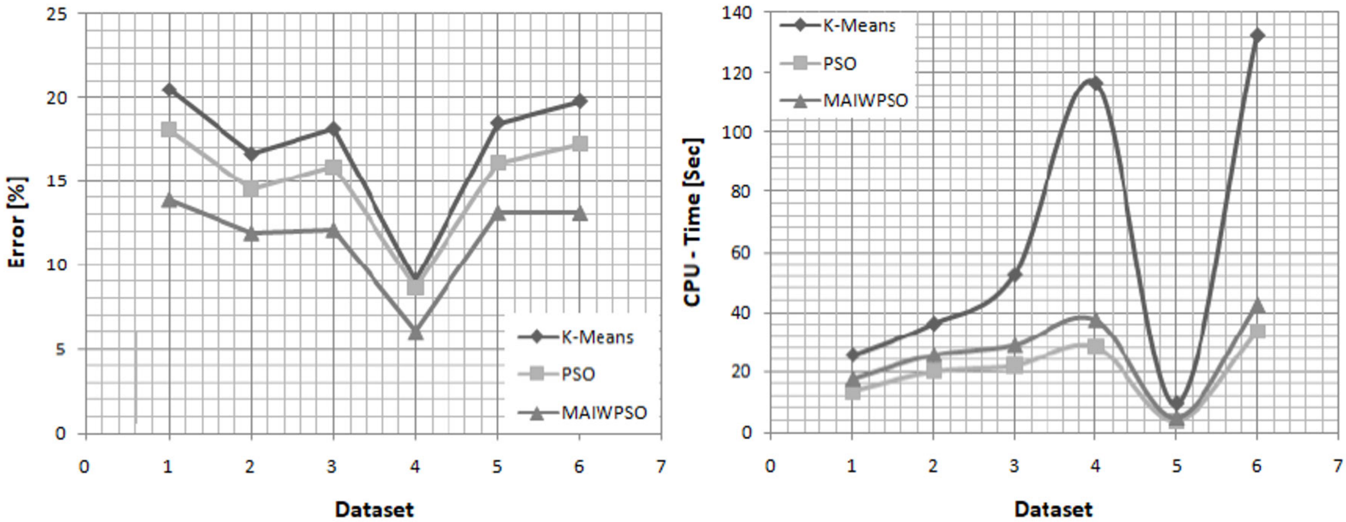
Table 3 Result of PSO for clustering problems

Sr. no.	*Test problem*	*Number of runs*					*Mean*	*% Error*	*Time (sec)*
		1	*2*	*3*	*4*	*5*			
1	Iris	96.68	103.14	105.34	98.89	104.35	101.68	18.03	13.38
2	Thyroid	1,859.87	1,878.76	1,907.87	1,863.72	1,895.43	1,881.13	14.49	19.89
3	Heart	1,642.71	1,632.10	1,639.82	1,625.01	1,624.71	1,632.87	15.84	21.87
4	Breast cancer	3,129.68	3,137.81	3,242.56	3,195.78	3,118.32	3,164.83	8.60	28.44
5	Crude oil	277.79	277.11	277.15	277.03	277.32	277.28	16.11	3.63
6	Pima	2,785.03	2,772.02	2,777.81	2,761.71	2,781.68	2,775.65	17.21	33.63

Table 4 Result of K-means for clustering problems

Sr. no.	*Test problem*	*Number of runs*					*Mean*	*% Error*	*Time (sec)*
		1	*2*	*3*	*4*	*5*			
1	Iris	94.67	92.89	94.47	91.63	91.69	93.07	13.86	17.39
2	Thyroid	1,683.62	1,619.17	1,673.87	1,676.29	1,685.15	1,667.62	11.89	25.46
3	Heart	1,508.39	1,516.11	1,506.72	1,514.13	1,512.10	1,511.49	12.08	28.87
4	Breast cancer	2,799.77	2,832.19	2,830.23	2,822.96	2,806.65	2,818.36	6.01	28.87
5	Crude oil	226.29	227.09	226.77	227.98	227.77	227.18	13.10	4.79
6	Pima	2,711.61	2,724.17	2,705.19	2,699.20	2,698.23	2,707.68	13.08	42.37

Figure 3 Mean error rate and mean computational time comparison, (a) mean error rate (b) mean computational time



(a) (b)

6 Conclusions

Swarm based optimisation algorithms are effective tools to perform high quality clustering. In current paper we proposed a modified adaptive inertia weight PSO (MAIWPSO) for data clustering to obtain high quality clusters centre in comparison of PSO and K-means. New MAIWPSO use particle fitness value in computation of inertia weight a maintain the swarm diversity during the course of searching. K-means, PSO and MAIWPSO were simulated on 6 standard dataset namely iris, thyroid, heart, breast cancer, crude oil and pima. Outcomes of experiments have been given in Tables 2 to 4. These results depict that MAIWPSO is better technique for clustering compared to PSO and K-means over criterion clustering metric (sum of intra cluster distance), error rate and computational time. Thus MAIWPSO is efficient, robust, easy to tune and produced high quality solutions compared to PSO and K-means. As future work we suggest to test the performance of MAIWPSO for clustering with other swarm based optimisation techniques like glowworm swarm optimisation, cuckoo search, gravitational search algorithm, grey wolf optimiser differential evolution, artificial bee colony, ant colony on different datasets and discover the best one clustering technique over given criterion.

References

Alswaitti, M., Albughdadi, M. and Isa, N.A.M. (2018) 'Density-based particle swarm optimization algorithm for data clustering', *Expert Systems with Applications*, Vol. 91, No. C, pp.170–186.

Bandyopadhyay, S. and Maulik, U. (2002) 'An evolutionary technique based on K-means algorithm for optimal clustering in RN', *Information Sciences*, Vol. 146, Nos. 1–4, pp.221–237.

Bouyer, A. and Hatamlou, A. (2018) 'An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms', *Applied Soft Computing*, Vol. 67, No. C, pp.172–182.

Chen, C.Y. and Ye, F. (2012) 'Particle swarm optimization algorithm and its application to clustering analysis', in *Electrical Power Distribution Networks (EPDC), 2012 Proceedings of 17th Conference on*, May, pp.789–794, IEEE.

Eberhart, R. and Kennedy, J. (1995) 'A new optimizer using particle swarm theory', in *Micro Machine and Human Science, 1995. MHS'95, Proceedings of the Sixth International Symposium on*, October, pp.39–43, IEEE.

Esmin, A.A. and Matwin, S. (2012) 'Data clustering using hybrid particle swarm optimization', in *International Conference on Intelligent Data Engineering and Automated Learning*, August, pp.159–166, Springer, Berlin, Heidelberg.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., Foufou, S. and Bouras, A. (2014) 'A survey of clustering algorithms for big data: taxonomy and empirical analysis', *IEEE Transactions on Emerging Topics in Computing*, Vol. 2, No. 3, pp.267–279.

Ghali, N.I., El-Dessouki, N., Mervat, A.N. and Bakrawi, L. (2009) 'Exponential particle swarm optimization approach for improving data clustering', *International Journal of Electrical, Computer, and Systems Engineering*, Vol. 3, No. 4, pp.208–212.

Gupta, I., Yadav, V., et al. (2019) 'Medical Data clustering based on particle swarm optimization and genetic algorithm', *International Journal of Advanced Intelligence Paradigms (IJAIP)*, November, Vol. 14, Nos. 3/4, pp.345–358, Inderscience Publication, ISSN online: 1755-0394 ISSN print: 1755-0386.

Izakian, H., Abraham, A. and Snel, V. (2009) 'Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization', in *Nature & Biologically Inspired Computing, NaBIC 2009, World Congress on*, December, pp.1690–1694, IEEE.

Izakian, Z., Mesgari, M.S. and Abraham, A. (2016) 'Automated clustering of trajectory data using a particle swarm optimization', *Computers, Environment and Urban Systems*, Vol. 55, pp.55–65.

Jaskowiak, P.A., Campello, R.J. and Costa Filho, I.G. (2013) 'Proximity measures for clustering gene expression microarray data: a validation methodology and a comparative analysis', *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, Vol. 10, No. 4, pp.845–857.

Kao, I.W., Tsai, C.Y. and Wang, Y.C. (2007) 'An effective particle swarm optimization method for data clustering', in *Industrial Engineering and Engineering Management, 2007 IEEE International Conference on*, December, pp.548–552, IEEE.

Kao, Y.T., Zahara, E. and Kao, I.W. (2008) 'A hybridized approach to data clustering', *Expert Systems with Applications*, Vol. 34, No. 3, pp.1754–1762.

Li, H., He, H. and Wen, Y. (2015) 'Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation', *Optik-International Journal for Light and Electron Optics*, Vol. 126, No. 24, pp.4817–4822.

Lin, C.R., Liu, K.H. and Chen, M.S. (2005) 'Dual clustering: integrating data clustering over optimization and constraint domains', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 5, pp.628–637.

Omran, M.G., Salman, A. and Engelbrecht, A.P. (2006) 'Dynamic clustering using particle swarm optimization with application in image segmentation', *Pattern Analysis and Applications*, Vol. 8, No. 4, p.332.

Otto, C., Wang, D. and Jain, A.K. (2018) 'Clustering millions of faces by identity', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 2, pp.289–303.

Parveen, H., Yadav, V. et al. (2018) 'A particle swarm optimization strategy using QSAR modeling on the second generation neural network', *IEEE International Conference on Advances in Computing, Communication Control & Networking (ICACCCN-2018)*, October, pp.1188–1199, ISBN: 978-1-5386-4119-4.

Saha, S. and Das, R. (2018) 'Exploring differential evolution and particle swarm optimization to develop some symmetry-based automatic clustering techniques: application to gene clustering', *Neural Computing and Applications*, Vol. 30, No. 3, pp.735–757.

Sengupta, S., Basak, S. and Peters, R.A. (2018) 'Data clustering using a hybrid of fuzzy C-means and quantum-behaved particle swarm optimization', in *Computing and Communication Workshop and Conference (CCWC), 2018 IEEE 8th Annual*, January, pp.137–142, IEEE.

Shi, Y. and Eberhart, R. (1998) 'A modified particle swarm optimizer', in *Evolutionary Computation Proceedings, IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference on*, May, pp.69–73, IEEE.

Tsai, C-Y. and Kao, I-W. (2011) 'Particle swarm optimization with selective particle regeneration for data clustering', *Expert Systems with Applications*, Vol. 38, No. 6, pp.6565–6576.

Zhan, Z.H., Zhang, J., Li, Y. and Chung, H.S.H. (2009) 'Adaptive particle swarm optimization', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 39, No. 6, pp.1362–1381.