



A Novel Approach of Order Clustering using Combination of Particle Swarm Optimization and Genetic Algorithm

Jyoti Sinha¹, Mr. Deepak Shrivastava²

¹M.Tech. Scholar, CSE Department, DIMAT, Raipur

²Asst.Professor, CSE Department, DIMAT, Raipur

Abstract-

To extract some useful and meaningful information from the large volume of dataset is known as learning. Learning can be classified as supervised and non-supervised learning. Clustering comes under the category of the unsupervised. Genetic algorithm (GA) and Particle swarm optimization (PSO) are two well-known method of clustering. Premature convergence is most common problem associated with the PSO. This paper address this problem and suggest a method to overcome this problem of PSO for using in order clustering. Simulation results shows that he proposed method outperform the existing methods.

Keywords- Clustering, Supervised Learning, Unsupervised Learning, Hierarchical clustering

I. INTRODUCTION

In so many industries, production process consist of two phases, first phase is used for preparing the material and second phase is used for binding the material before line change occur. This process is very time consuming. If the labours in the production system fail to finish the task before the next order start, it produce idle time for the very expensive machine installed in the factory for this very purpose. More idle time creates less production. This time of problem is faced in so many industries and the only solution to this problem is to arrange more material and putting the increased number of labour in the task. Another solution of this problem is to cluster the order before scheduling process and then performing the scheduling of the similar order for reducing the material binding time. This is where order clustering comes in to the picture in industries.

Clustering can be considered as the most crucial unsupervised learning technique or classification method. Algorithm based on the clustering can be applied in wide range of application, some of which are data mining [1], Pattern Recognition [2] data compression algorithm [3], machine learning application etc. In case of clustering based problem, if the number of cluster k is known a priori, then clustering can be formulated by arranging all the available data in k groups in such a way that elements in the same cluster share some common properties and completely different from the elements of the other cluster. This is achieved by minimizing the some predefined optimization criterion. K-means clustering is one of the most widely used algorithm for clustering the data[5]. K-means clustering method, start the clustering process by initializing k -random centres and then divide the data in these different cluster by minimizing the criterion such as Euclidean distance. However k -means clustering has some serious drawbacks like trapping in local minima or local maxima and sensitiveness towards the initial cluster centres. One solution to this problem of the k -means- clustering is to hybridized it with some optimization method. Various optimaization algorithms such ant colony Optimization (ACO)[6], Bee colony optimization [7] etc, have been proposed in past.

Genetic algorithm [8], combination of genetic algorithm and k-means[9], pure k-mean[10], combination of self organizing map (SOM) and PSO[11,16,17,20], clustering based on PSO[12], different variants of k-means[13,14,15,18,19,22] are some noteworthy contribution in the field of optimization.

In this paper, authors suggest anew algorithm for order clustering. This new algorithm is a combination of particle swarm optimization and genetic algorithm. In this algorithm genetic algorithm is used in mutation process to generate new chromosomes which stops the premature convergence problem which is common in PSO along with add diversity in the solution.

II. METHODOLOGY

In this paper, a new method of clustering which is called HPSO(hybrid particle swarm optimization) have been proposed which is combination of the particle swarm optimization and genetic algorithm. In order to evaluate the efficiency and accuracy of the proposed method over the existing method, a performance comparision has also been done for all the three methods.

Block diagram of ovel all methodology is shown in the figure 2. First of all, four different dataset have been chosen for evaluating the performance. These data set are clustered using all the three methods. Clustered formed by all the three methods is then compared with the help of some performance measurement. On the basis of the results obtained , conclusion has been drawn. To cluster the dataset with the help of the Genetic Algorithm (GA), a program for genetic algorithm has been designed. Steps of the algorithms for the genetic algorithm is given in the next section.

A. Algorithm Steps of Genetic Algorithm

Following are the algorithm steps for Genetic algorithm applied in this project work-

Step-1 Initialize the population of n chromosomes randomly (Probable solution of the problem)

Step-2 Calculate the fitness value or score for each chromosome x by using fitness function $f(x)$.

Step-3 Generate the new population by repeating the below mentioned steps until the completion of new population.

- i. **Selection process**- select the chromosomes from a population as per their fitness value (Higher the fitness score better the chances of selection).
- ii. **Crossover Process**- Crossover the parent to generate a new offspring(children) as per the given crossover probability. Without crossover, offspring are exact replica of parents.
- iii. **Mutation process**- Mutate two offspring with mutation probability.
- iv. **Accepting Process**- Place a new offspring in population to create new population.

Step-4 **Replacement Process**- For further processing, use the newly generated population.

Step-5 **Test Process**- If the stopping condition met then stop and return the best solution in current population else (i.e. if stopping condition not met) then go to step 2.

Similarly a program is designed for particle swarm optimization whose algorithm steps are given in the next section.

B. Algorithm Steps For Particle Swarm Optimization Algorithm (PSOA)

Following are the algorithm steps for particle swarm optimization-

Step-1 First of all Initialize the population.

Step-2 for each particle computes the fitness score or fitness value.

Step-3 if the current fitness value is better than pBest then assign the current fitness value as pBest else keeps the pBest value intact.

Step-4 Assign the best particle pBest value as the gBest.

Step-5 Compute the velocity of each particle.

Step-6 Update the data value of each particle on the basis of velocity values.

Step-7 if either set target or maximum epoch is reached then stop else go to step 2.

In the next phase of this project, a program for the proposed HPSO (Hybrid Particle Swarm Optimization) algorithm has been implemented with the help of steps given in the next section.

C. Proposed Hybrid particle Swarm Optimization (HPSO)

Past research work performed on the efficiency of the PSO reveals that in PSO, selection of the parameter is very difficult and uphill task. It has also been found by comparing the performance of the Genetic algorithm with the PSO algorithm that PSO perform very well during the early iteration but as we get closer to the optimal solution, PSO starts producing problems. Apart from this, Updating of the velocity of the particle play very important role for deciding the PSO behaviour. Some it has been seen that after attaining the global best position, particle start drifting away from this point provided the previous velocity of the particle or inertia weight is non zero.

Some if the particle velocity is equal to zero or nearly equal to zero then particle will not change its position resulting in a convergence problems.

This work is an attempt to address this problem. This problem can be overcome by designing a new algorithm called HPSO. In this new approach, mutation process of the GA is included in this algorithm. This is performed by mutating the search space dynamically

This step causes the stagnation process to stop and incorporate the population diversity.

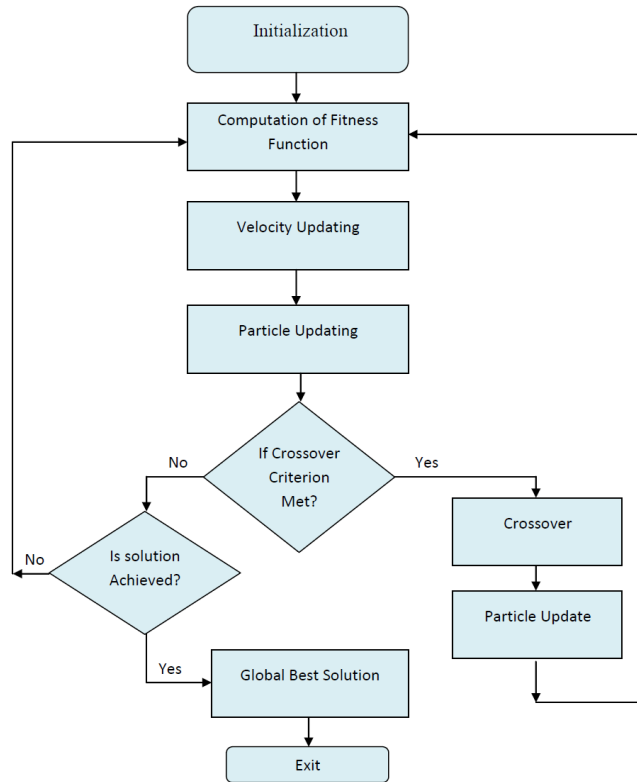


Figure 3 Flowchart of Proposed Hybrid Particle Swarm Optimization(HPSO)

This step help the swarm to get away from the local minima and extend its circle to different zone of the search space. Due to this , the proposed algorithm acquire the ability to balance the global and local search to get the better convergence.

Algorithm start by randomly choosing the swarm particle and making it move to different direction in the search space. Mutation operation is performed after this. Block diagram of the proposed HSOM method is shown in the figure 4.1.

In the proposed method , first of all, population is randomly generated. Elements or individuals of this population is regarded as the particle. In the next step of the algorithm , fitness function is computed. Velocity of each particle is then calculated or updated. In the next step position of the particle is also updated. Then crossover is introduced here i.e. genetic algorithm. Crossover operation is done to produce new elements in search space. Particle is again updated after this crossover. This procedure is repeated till the cross over criterion is satisfied. If does not satisfy then we check whether the solution goal is achieved or not. If solution goal is achieved then we declare it as the global solution. If solution goal is not achieved then again fitness function is computed and so on.

D. Algorithm Steps for the Proposed method

Step by step algorithm for the proposed method for clustering the data of database is explained in the next section-

Step-1 Initialized the required parameter i.e. size of population, learning factor $c1$ and $c2$, maximum velocity V_{max} , inertia weight W etc.

Step-2 Initialize the initial position of each particle randomly, velocity V_{id} and X_{id} .

Step-3 For each particle compute the fitness value for suitability adjustment

$$\text{Fitness value} = \sum_{j=1}^k \sum_{x \in c_{ij}} \|x - z_{ij}\|$$

Here

x represent the data elements of the data vector to be clustered. k represent cluster number. c_{ij} Represent the no. of data vectors of i^{th} particle in cluster j

$\|x - z_{ij}\|$ Represent Euclidean distance between data vector and cluster centroids

Step-3a Compute the Euclidean distance $d(x, z_{ij})$ between each data Elements x with data cluster z_{ij} .

$$d(x, z_{ij}) = \|x - z_{ij}\|$$

Step-3b Assign data element x to the cluster z_{ij} according to the rule given below-

$$d(x, z_{ij}) = \min \{d(x, z_{i1}), \dots, d(x, z_{ik})\}$$

Step-3c Compute the fitness as given in step 3.

Step-4 Apply genetic algorithm (reproduction, crossover, mutation) to $2N$ population to create $2N$ population.

Step-4a Apply reproduction process i.e. choose the $2N^s$ best chromosome according to the fitness score.

Step-4b Apply Crossover process for updating the $2N$ individual.

Step-4c Apply Mutation Process with 5% mutation probability to $2N$ Chromosome as per the equation given below

$$x' = x + \text{rand} \times N(0, 1)$$

Step-5 Compute the fitness value of the $2N$ individuals and rank them According to their fitness values.

Step-6 Updation of the global best (p_{gd}) and local best (p_{id}).

Step-7 Apply crossover to the global best and local best to get two child particles and choose the child having smaller fitness value as the final child particle for further processing.

Step- 8 (a) Do velocity updation as per the following equation

$$V_{id}^{\text{new}} = W * V_{id}^{\text{old}} + c1 * \text{rand1} * (P_{id} - X_{id}) \\ + c2 * \text{rand2} * (P_{gd} - X_{id})$$

Here c_1 and c_2 are two positive constant And rand_1 and rand_2 are two random function within the range[0,1]. Inertia weight is represented by W .

Step-8 (b) Update the V_{id} using mutation as per the following equation

$$V_{old}^{new} = V_{old}^{new} + \text{rand} \times N(0,1)$$

Step-9 (a) do the position vector X_{id} updation of top N particles by following formula

$$X_{id}^{new} = X_{id}^{old} + V_{id}^{new}$$

Step-9 (b) Update the X_{id} using mutation as per the following equation

$$X_{old}^{new} = X_{old}^{new} + \text{rand} \times N(0,1)$$

Step-11 If stopping criterion meet then go to step 12 else go to step 3.

Step-12 Particle with the minimum fitness value in last generation is the solution

III. EXPERIMENTAL RESULTS

For evaluating the performance of the all the three methods of clustering, MSE and SED are used for performance measure. Mean square error (MSE) can be computed using the following formula-

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

In this formula $\hat{Y}_i$ predicted value is represented by the $\hat{Y}_i$ and observed value is represented by the Y_i .

Sum of Euclidean distance (SED) is computed using the following formula

$$SED = \sum_{i=1}^n \sum_{j=1, x_i \in C_j}^k d(x_i, m_j)$$

n =total number of data

$C_j = j^{\text{th}}$ cluster

K =number of cluster

D =Euclidean Distance

x_i = dimension of i^{th} data.

M_j =coordinate of the j^{th} cluster's center.

First of all, all the database are clustered by using GA, PSO and HPSO method. MSE for different number of generations are computed and the graph between MSE and Number of generations are plotted. This graph is shown in the figure 5. Number of generations is taken along the x-axis and the value of MSE is taken along the y-axis. From the graph shown in the figure 5, it is clear that as we increase the number of generations, MSE value starts decreasing for all the three method but for HPSO it is

found to be lowest in all the number of generations. In the graph 5 MSE curve for the PSO method is represented by the green line, red line represent the graph for genetic algorithm(GA) and blue color line represent the MSE curve for the HPSO method. Lowest value of the MSE for HPSO method represent that the cluster formed by this method is best in term of accuracy.

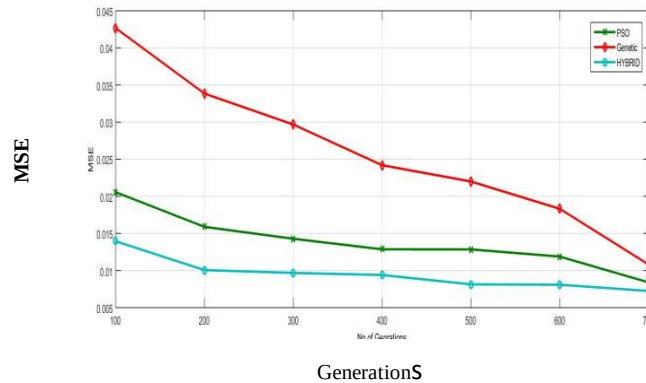


Figure 5.1 MSE comparison Graph For all the three algorithm

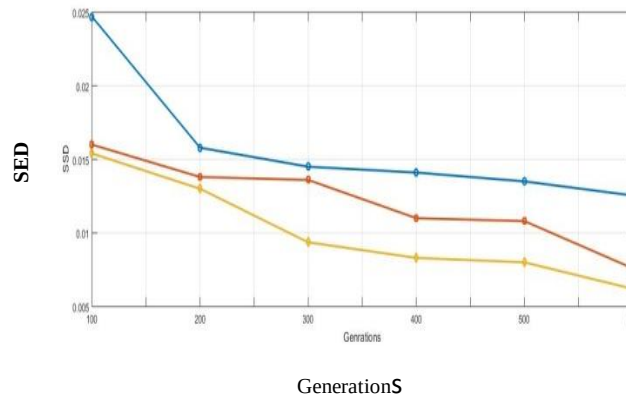


Figure 5.2 SED comparison Graph For all the three algorithm

SED(Sum of Euclidean Distance) is another method of evaluating the performance of the clustering method. SED curve for all the three methods for different number of generations is shown in the figure 6. Here blue line represent the SED curve for the PSO method, Orange color line represent the SED curve for the GA method and yellow line represent the SED curve for the HPSO method. From this graph it is evident that as we increase the number of generations, SED value start decreasing. SED value is found to be lowest for HPSO method for all the number of generations.

IV. CONCLUSION

Order clustering is one of the method of reducing the material binding time in production system. In this paper, hybrid particle swarm optimization (HPSO) method is used for order clustering purpose.

Though we can use genetic algorithm(GA) and PSO(particle swarm optimization) for order clustering but for getting better efficiency and accuracy of the clustering, in this paper we have proposed a new method which is called. For evaluating the performance of all the three method of clustering we have computed MSE (Mean square error)and SED. A simulation program

is designed and simulation is carried out by applying GA, PSO and HPSO algorithm in all the four dataset i.e. Glass, wine, Iris and vowel. Both the parameters are found to be lowest for HPSO algorithm which shows the better performance of this method as compared to the other methods.

References

- [01] C. Pizzuti and D. Talia, "P-AutoClass: scalable parallel clustering for mining large data sets", in *IEEE transaction on Knowledge and data engineering*, Vol. 15, pp. 629-641, May 2003.
- [02] K. C. Wong and G. C. L. Li, "Simultaneous Pattern and Data Clustering for Pattern Cluster Analysis", in *IEEE Transaction on Knowledge and Data Engineering*, Vol. 20, pp. 911-923, Los Angeles, USA, June 2008.
- [03] J. Marr, "Comparison Of Several Clustering Algorithms for Data Rate Compression of LPC Parameters", in *IEEE International Conference on Acoustics Speech, and Signal Processing*, Vol. 6, pp. 964-966, January 2003.
- [04] X. L. Yang, Q. Song and W. B. Zhang, "Kernel-based Deterministic Annealing Algorithm For Data Clustering", in *IEEE Proceedings on Vision, Image and Signal Processing*, Vol.153, pp. 557-568, March 2007.
- [05] Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley & Sons.
- [06] M. Darigo, M. Birattari, T. Stutzle, "Ant Colony Optimization". In: *IEEE Computational Intelligent Magazine*, Vol. 1, pp. 28-39, 2006.
- [07] D.T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, M. Zaidi, "The Bees Algorithm - A Novel Tool for Complex Optimisation Problems". *Proceedings of IPROMS 2006 Conference*, pp. 454-461, 2006.
- [08] J. Holland, *Adaptation in Natural and Artificial System*, University of Michigan Press, Ann Arbor, MI, 1975.
- [09] C.A. Murthy, N. Chowdhury, In search of optimal clusters using genetic algorithms, *Pattern Recognition Letters* 17 (8) (1996) 825-832.
- [10] D. Pelleg, A. Moore, X-means: extending K-means with efficient estimation of the number of clusters, *Proceedings of the 17th International Conference on Machine Learning*, Morgan Kaufmann, San Francisco, CA, 2000 pp. 727-734.
- [11] X. Xiao, E.R. Dow, R. Eberhart, Z.B. Miled, R.J. Oppelt, Gene clustering using self organizing maps and particle swarm optimization, *Proceedings of the International Parallel and Distributed Processing Symposium*, 2003, pp. 22-28.
- [12] D.W. van der Merwe, A.P. Engelbrecht, Data clustering using particle swarm optimization, *The Congress on Evolutionary Computation* (2003) 215-220.
- [13] Xiao, X., Dow, E. R., Eberhart, R., Miled, Z. B., & Oppelt, R. J. (2003, April). Gene clustering using self-organizing maps and particle swarm optimization. In *Parallel and Distributed Processing Symposium, 2003. Proceedings. International* (pp. 10-pp). IEEE.
- [14] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, and S. Brown (2004), 'FGKA: A Fast Genetic K-means Clustering Algorithm', *ACM* 1-58113-812-1
- [15] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, and S. Brown (2004), 'Incremental genetic K-means algorithm and its application in gene expression data analysis', *BMC Bioinformatics* 5:172.
- [16] Cui, X., Potok, T. E., & Palathingal, P. (2005, June). Document clustering using particle swarm optimization. In *Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE* (pp. 185-191). IEEE.
- [17] M.G.H. Omran, A. Salman, A.P. Engelbrecht, Dynamic clustering using particle swarm optimization with application in image segmentation, *Pattern Analysis and Applications* 8 (2006) 332-344.
- [18] H. Abolhassani, M. Mahdavi, 2009. Harmony K-means algorithm for document clustering. *Data Mining Knowledge Discovery*.18:370-391.
- [19] Dharmendra K Roy and Lokesh K Sharma, 2010. Genetic K-Means clustering Algorithm for mixed numeric and categorical data. *International journal of Artificial Intelligence & Applications* Vol1 No2.
- [20] Feng, L., Qiu, M. H., Wang, Y. X., Xiang, Q. L., Yang, Y. F., & Liu, K. (2010). A fast divisive clustering algorithm using an improved discrete particle swarm optimizer. *Pattern Recognition Letters*, 31(11), 1216-1225.

- [21] Senthilnath, J., Omkar, S. N., & Mani, V. (2011). Clustering using firefly algorithm: performance study. *Swarm and Evolutionary Computation*, 1(3), 164-171.
- [22] Mahendiran, A., Saravanan, N., Subramanian, N. V., & Sairam, N. (2012). Implementation of K-means clustering in cloud computing environment. *Research Journal of Applied Sciences, Engineering and Technology*, 4(10), 1391-1394.