

PSO BASED MULTIDIMENSIONAL DATA CLUSTERING: A SURVEY

Satyabrata Sahoo¹, Dr. Anjan Ghosh²

¹(Electrical Engineering, Gandhi Engineering College, Bhubaneswar)

²(Electrical Engineering, Gandhi Engineering College, Bhubaneswar)

Abstract

Data clustering is considered as one of the most promising data analysis methods in data mining and on the other side K- Means is the well known partitional clustering technique. Nevertheless, K-Means and other partitional clustering techniques struggle with some challenges where dimension is the core concern. The different challenges associated with clustering techniques are preknowledge of initial centers of clusters, problem of stagnation, multiple cluster membership problem, dead unit problem, and slow or premature convergence to local search space. So as to resolve these clustering limitations, an eminent choice is to adapt the Swarm Intelligence (SI) inspired optimization algorithms. This paper presents an overview of the research on an applicability of different Particle Swarm Optimization (PSO) variants for clustering multidimensional data along with the basic concepts of PSO as well as data clustering. It also puts forward an idea of new and advance PSO variant in order to deal with multidimensional dataclustering.

Keywords

BRAPSO, Data Clustering, Particle Swarm Optimization (PSO), Subtractive Clustering (SC), Swarm Intelligence (SI).

I. INTRODUCTION

Data mining has been proved as one of the favorite areas of researchers with its absolutely outstanding applications in variety of fields, exploring new approaches to visualize, examine, and uncover any hidden patterns in data automatically. Data mining system can be defined as an integration of techniques from multiple disciplines namely statistics, database & data warehouse, information retrieval, machine learning, pattern recognition, neural networks, image & signal processing, and artificial intelligence etc. [24], whereas, Artificial intelligence (AI) is a technology and a branch of computer science that studies and develops intelligent machines and software. Several research textbooks have defined Artificial Intelligence as the concept of studying and designing the intelligent agents. These intelligent agents are nothing but a system that perceives its environment and tries to achieve maximum chances of success. AI research involves different technical issues and several approaches are struggling to solve them. Some subfields are aimed to find solution of specific problems whereas others focus on one of several possible approaches. By using AI systems, computer can make decisions without any human interference. It actually imparts a sixth sense to the different technologies for exploring new data. AI is applied to data mining systems so as to find a new way to explore the meaningful information that can be useful for several future decisions making processes.

For this research work, these two renowned domains are chosen in order to utilize the benefits of both fields and apply it to some real world applications. The well admired data analysis method in data mining system is nothing but data clustering itself. It finds its applications in biological sequences, scheduling, collective behavior, image segmentation, biomedical, graphics etc. It is the method of grouping similar data objects with maximum inter-cluster distance and minimum intra-cluster distance. Compact clustering is essential for grouping similar items together to provide accurate information so that meaningful information can be retrieved from it and can be used for decision making purpose. There is variety of approaches for to derive good results of clustering, even though they suffer from several challenging drawbacks. Hence the researchers found data clustering as crucial area of research to enhance the results.

Applicability of AI to solve these clustering challenges is gaining importance these days. Swarm Intelligence (SI) is derived as a relatively new subfield of AI which studies the emergent collective intelligence of groups of simple agents. It adapts the social behavior that can be observed in nature elements, such as, ant colonies, birds flock, bee hives, and fish schools etc. In these cases, number of individuals with limited capabilities produces an intelligent solution for complex problems. The different swarm intelligence inspired optimization algorithms are Ant Colony Optimization (ACO), Bee Colony Optimization (BCO), Fish School Optimization (FSO), Particle Swarm Optimization (PSO), etc. Amongst these optimization techniques, PSO and ACO are mostly used [21].

PSO is one of the SI paradigms that have received widespread attention in research these days. It is a novel population based stochastic search algorithm that provides a solution to complex non-linear optimization problem. It is an evolutionary computation technique which simulates the movement and flocking of birds that performs a global search in the solution space. PSO produces better results in complicated and multi-peak problems with few parameters to adjust giving fast as well as accurate computation results which makes it as popular optimization technique in swarm intelligence field [26]. Hence, PSO will be a great choice for finding solution to the most challenging issues in clustering. For a brief idea about the concerned area of work, see Figure 1.

The rest of the paper is organized as follows. In Section II, a brief literature survey on PSO and its different variants is described. Section III provides an analysis of clustering techniques and related challenges. The basic concept of PSO algorithm, along with the scope of research that motivates the future work is described in Section IV. Section V analyses the characteristics of different data sets having multiple dimensions and proposal of new variant of PSO for handling issues while clustering multidimensional data, followed by Section VI, which concludes this paper providing future direction.

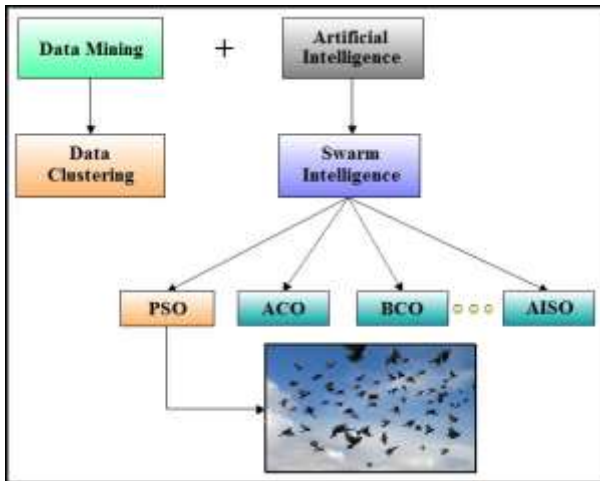


Fig 1: Illustration of concerned area of work

II. LITERATURE SURVEY

S. Rana et al. [1] proposed a new improved algorithm named as Boundary Restricted Adaptive Particle Swarm Optimization (BRAPSO) algorithm with boundary restriction strategy for particles that travel outside the boundary search space during PSO process. Nine data sets were used for the experimental testing of BR-APSO algorithm, and its results were compared with PSO as well as some other PSO variants namely, K-PSO, NM-PSO, and K-Means clustering algorithms. It has been found that the proposed algorithm is robust, generates more accurate results and its convergence speed is also fast as compared to other algorithms.

Mariam El-Tarabily et al. [2] presented a hybrid Particle Swarm Optimization, i.e. Subtractive + (PSO) clustering algorithm that performs fast clustering. They have tested the Subtractive + (PSO) clustering algorithm as well as the Subtractive clustering algorithms and PSO on three different data sets for comparison purpose. From the results, it is proved that the proposed clustering algorithm provides better clustering results than the other algorithms.

Sandeep Rana et al. [3] proposed a new Hybrid Sequential clustering approach. They have used PSO and K-Means algorithm in sequence for data clustering. This approach was proposed to overcome the drawbacks of both algorithms as well as improves clustering and avoids being stagnated. Four kinds of data sets have been tested in order to obtain comparative results. For comparison purpose, different algorithms such as PSO, K-Means, Hybrid K-Means PSO, and Hybrid K-Means + Genetic Algorithm were considered. The proposed algorithm generates more accurate and robust clustering results.

Mehdi Neshat et al. [4] developed a cooperative algorithm based on PSO and K-Means algorithm. The proposed algorithm takes benefits of both global search ability of PSO as well as local search ability of K-Means. When tested and compared with PSO; K-Means, KPSO hybrid algorithm, and Contraction Factor PSO (CF-PSO) on six data sets, it has been shown that the proposed algorithm has satisfactory efficiency and robustness.

Serkan Kiranyaz et al. [5] invented two novel techniques. The first one is Multi Dimensional-PSO (MD-PSO), which is proposed to ignore the mandatory condition of knowing a fixed dimension *a priori*, which is a common drawback of different swarm optimizers. Even so, MD PSO is vulnerable to premature convergences because of lack of divergence. To address this problem, the second technique i.e. Fractional Global Best Formation (FGBF) technique is proposed. These two techniques have been tested for nonlinear function minimization and data clustering, which are considered applications for experiment. After a large number of testing phases, it came to know that in both application domains, MD PSO with FGBF exhibits a remarkable speed gain and converges to the global optima at the true dimension regardless of the dimension of the search space, size of swarms, and the problem complexity.

Bara'a Ali Attea et al. [6] discovered that performance of clustering algorithms degrades with more and more overlaps among clusters in a data set. These facts have motivated to develop a fuzzy multi-objective particle swarm optimization framework (FMOPSO) in an innovative fashion for data clustering, which is able to deliver more effective results than state-of-the-art clustering algorithms. To ascertain the superiority of the proposed algorithm, number of statistical tests has been carried out on a variety of numerical and categorical real life data sets.

Suresh Chandra Satapathy et al. [7] introduced an idea of an algorithm that can combine dimensionality reduction techniques of weighted PCs with AUTO-PSO for clustering. The intention behind it was to reduce complexity of data sets and speed up the Auto PSO clustering process. A significant improvement in total runtime has been achieved. Moreover, the clustering accuracy of the dimensionality reduction technique i.e. AUTO-PSO clustering algorithm is comparable to the one that uses full dimensionspace.

Li-Yeh Chuang et al. [8] invented an improved particle swarm optimization based on Gauss chaotic map for clustering. Gauss chaotic map provides the significant chaos distribution to balance the exploration and exploitation capability for search process. This easy and fast function generates a random seed processes, and further improve the performance of PSO due to their unpredictability. In the experimental analysis, the eight different clustering algorithms were compared on six test data sets. The results indicated that the performance of the proposed one is appreciably better than the performance of other existing algorithms.

Xiaohui Cui et al. [9] presented a Particle Swarm Optimization (PSO) document clustering algorithm. Unlike, localized searching of the K-Means algorithm, PSO clustering algorithm used to perform a globalized search in the entire solution space. In the experiments conducted, they have applied the K-Means PSO, and hybrid PSO clustering algorithm on four different text document data sets. From the comparative results, the hybrid PSO algorithm can generate more compact clustering results than the K-Means algorithm.

R.Karthi et al. [10] discovered fascinating features of well known PSO algorithms for solving continuous function optimization problems and evaluated by considering real world data clustering problems. To solve different data clustering problems, three performance clustering metrics such as TRace Within criteria (TRW), Variance Ratio Criteria (VRC) and Marriott Criteria (MC) are considered. The results obtained by the PSO variants were compared with the basic Genetic Algorithm, PSO algorithm, and Differential Evolution algorithms for comparison purpose. To study the convergence behavior of the PSO algorithms using run length distribution, a detailed performance analysis has been carried out.

Feng Xie et al. [11] worked out an adaptive particle swarm optimization (PSO) on individual level. By analyzing the social model of PSO, a replacing criterion based on the diversity of fitness between current particle and the best historical experience is introduced to maintain the social attribution of swarm adaptively by removing inactive particles. Three benchmark functions were tested which indicates its improvement in the average performance.

Jianchao Fan et al. [12] proposed a particle swarm optimization approach with dynamic neighborhood based on kernel fuzzy clustering and variable trust region methods (called FT-DNPSO) for large-scale optimization. It adaptively adjusts the initial region and clusters different dimension into groups, which expedites convergence and search in the effective range. The adaptive strategy avoids or alleviates the prematurity of the PSO algorithm. The simulation results, with eight classical benchmark functions, twenty CEC2010 test ones and soft computing special session test; demonstrate that the proposed FT-DNPSO outperformed other PSO algorithms for large-scale optimization.

K. Premalatha et al. [13] presented the hybrid approach of PSO with Genetic Algorithm (GA). The proposed hybrid PSO systems find a better solution without trapping in local maximum, and to achieve faster convergence rate. This is because when the stagnation of PSO occurs, GA diversifies the particle position even though the solution is worse. This makes PSO-GA more flexible and robust. Unlike standard PSO, PSO-GA is more reliable in giving better quality solutions with reasonable computational time. Experiment results are examined with benchmark functions and results show that the proposed hybrid models outperform the standard PSO.

Chetna Sethi et al. [14] proposed a Linear PCA based hybrid K-Means clustering and PSO algorithm (PCA-K-PSO). In (PCA-K-PSO) algorithm the fast convergence of K-Means algorithm and the global searching ability of Particle Swarm Optimization (PSO) are combined for clustering large data sets using Linear PCA. Better clustering results can be obtained with PCA-K-PSO as compared to ordinary PSO. This was effectively developed in order to make its use for efficient clustering of high-dimensional datasets.

Ahmed A. A. Esmin et al. [15] presented a literature survey on the PSO algorithm and its variants to clustering high dimensional data. An attempt is made to provide a guide for the researchers who are working in the area of PSO and high dimensional data clustering.

The comparative analysis of five best of different studied PSO based clustering techniques are briefly described in Table 1.

III. CLUSTERING

What is Data Clustering?

Data clustering is one of the most popular data analysis techniques in data mining [22]. It is a process of partitioning an unlabeled data set into groups, where each group contains objects which are similar to each other with respect to a certain similarity measure and different from those of other groups. The idea of data clustering is, whenever we are given with a large amount of data, a huge number of data need to be summarized into a small number of groups or categories in order to further facilitate its analysis [18]. Nevertheless, finding these groupings or trying to categorize the data is tough task unless the data is of low dimensionality (two or three dimensions at maximum). For this reason, some methods in soft computing have been proposed to solve such kind of problem whose basic aim is to maximize the inter cluster distance & minimize the intra cluster distance as depicted in the Figure 2. These methods are called *Data Clustering Methods*[25].

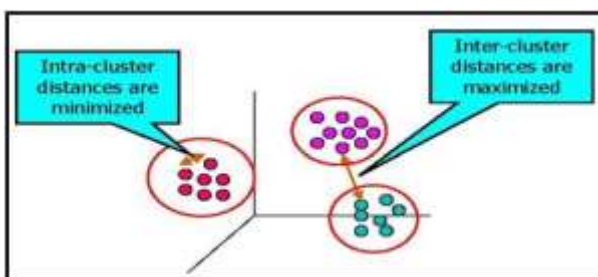


Fig 2: Clustering analysis process [1]

Clustering methods are majorly classified into two as follows:

Partitioning Methods: These methods partition the data set of size n into $(k \leq n)$ groups. They include well known algorithms like PAM, K-Means etc.

Hierarchical Methods: These are the methods that provide a tree view of clusters, called dendograms.

Various clustering techniques are studied [24][28], and analyzed classification of these techniques is shown in Figure 3.

Clustering Challenges

Clustering of data implies that the data set to be partitioned has to have an inherent grouping to some extent; otherwise for uniformly distributed data, it will fail finding clusters, or will lead to inaccurate partitions. Overlapping of data groups is another problem in clustering. Due to overlapping in groupings, sometimes it reduces the efficiency of the clustering method. This reduction is proportional to the amount of overlap between groupings. The knowledge of cluster centers apriori is very important for good quality clustering results. A cluster center specifies the heart of each cluster, hence with the help of similarity metric between the input vector and all the cluster centers, the system can find out the belongingness of given input vector to specific cluster, and determines the cluster which is the *nearest* or most similar one.

Some of the clustering techniques such as, K-Means and Fuzzy C-means clustering rely on knowing the number of clusters apriori. These are the algorithm which tries to partition the data into the given number of clusters. In other cases it is not necessary to have the number of clusters known from the beginning; instead the algorithm first finds the largest cluster, then second largest cluster, and so on. Examples of such type are Mountain and Subtractive clustering. In both cases a problem of known cluster numbers can be applied; however when we don't know the number of clusters in advance, we are not able to use K-Means and Fuzzy C-means clustering methods[23].

Table 1. Comparative analysis of five different PSO based clustering techniques

Paper Title/ Publication	Techniques	Data Sets/ Benchmark Functions	Purpose	PROS	Limitations	Future Work
A boundary restricted adaptive particle swarm optimization for data clustering. Springer 2012 S. Rana et al.[1]	BRAPSO	Art1, Art2, Vowel, Iris, Crude oil, CMC, Glass and Wine.	To use Boundary restriction strategy for compact clustering.	1. Handles particles that travel outside the boundary search space. 2. Fast convergence.	-	-
A PSO-Based Subtractive Data Clustering Algorithm. IJORCS 2013 Mariam El-Tarabily et al.[2]	Hybrid Subtractive clustering PSO	Iris, Wine, Yeast	To Estimate number of clusters and cluster center location.	1. Fast convergence and least fitness function. 2. Works well with multidimensional data without dimension reduction approach.	1. Multi-cluster membership problem.	1. Hybrid of subtractive clustering with any improved version of PSO.
A hybrid sequential approach for data clustering using K-Means and PSO algorithm. IJEST 2010 Sandeep Rana et.al.[3]	Hybrid Sequential Clustering Algorithm (K-Means + PSO)	Iris , Wine, Art1, Art2	To improve performance of hybrid PSO.	1. Overcomes the drawback of K-Means with improved error rate and good convergence speed. 2. Requires less iteration.	1. Based on random selection of initial cluster centers.	1. Variations in PSO algorithm and its hybridization with K-Means algorithm.
A New Cooperative Algorithm Based on PSO and K-Means for Data Clustering. 9 ~	Coop-KPSO K-Means, PSO	Iris, Pima, Wine, Glass, Sonar, WDBC	To utilize the advantages of K-Means & PSO and remove their weaknesses.	1. Efficient and robust clustering algorithm.	1. Don't work well with high dimensional data sets.	1. Increase local search ability around the best found position to improve result.

JCS 2012 Mehdi Neshat et al.[4]						
Fractional Particle Swarm Optimization in Multidimensional Search Space.	MD-PSO with FGBF	Sphere, DeJong, Rosenbrock, Rastrigin, Griewank, Schwefel, Giunta & synthetic data spaces	To address major problems with PSO with high dimension data.	1. Work well with multi dimensional datasets. 2. Able to have prior knowledge of the search space dimension and avoids premature convergence to local optima.	1. Occasional over-clustering can be encountered. 2. Computational cost increases with dimensions.	1. Other PSO variants can be used with MD-PSO or FGBF
IEEE Transaction April 2010 Serkan Kiranyaz et al.[5]						

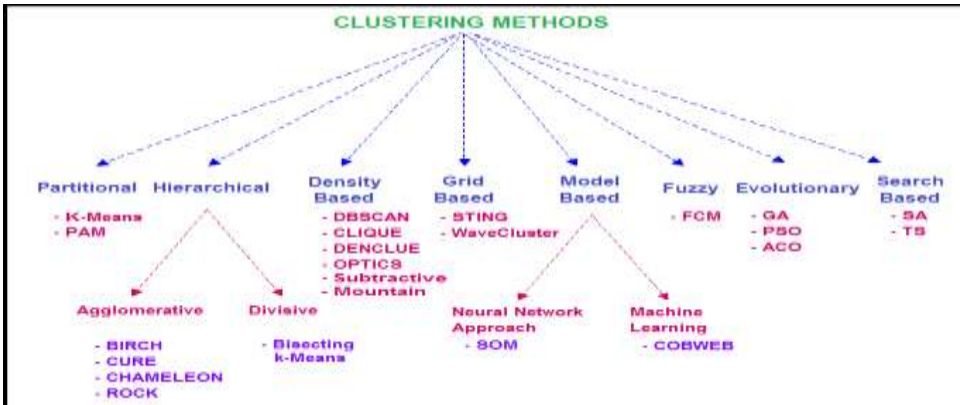


Fig 3: Analyzed classification of clustering techniques

After studying different clustering techniques we found four of them to be most promising but, these suffer with the limitations cum challenges as listed below:

- K-Means clustering produces higher accuracy and lower Root Mean Square Error (RMSE) than other techniques, and requires less computation time but the knowledge of the number of clusters a priori questions about the accuracy of clustering.
- Mountain clustering has a very poor performance regarding its requirement for huge number of computation and low accuracy. Mountain clustering is suitable only for problems with two or three dimensions.
- FCM produces similar results as K-Means clustering does, yet it requires more computation time than K-Means because of the fuzzy measures calculations involved in the algorithm.
- In subtractive clustering, value of the neighborhood radius r_n must be chosen carefully, since too small radii will result in neglecting the effect of neighboring data points, while large radii will result in a neighborhood of all the data points thus canceling the effect of the cluster.

All of the clustering techniques achieved enough high accuracy rates when tested on data sets. Although subtractive clustering presents greater impact because of highly compact clustering result with minimum RMSE and even works well with high dimensional data sets, but it is assumed that the problem data itself contains some overlapping in some of the dimensions; because of the high number of dimensions tend to disrupt the coupling of data and reduce the accuracy of clustering [2][23]. Hence, if subtractive clustering can be improved in order to deal with overlapping problem, even better results can be obtained. Thus, the above mentioned challenges open a new scope of research.

IV. PARTICLE SWARM OPTIMIZATION

Basic Model of PSO

Particle swarm optimization is a novel population-based stochastic search algorithm put forward originally by Dr. Russell C. Eberhart and Dr. James Kennedy in 1995. It is developed from swarm intelligence and is based on the research on bird and fish flock movement behavior. It provides a solution to complex non-linear optimization problem. In the basic particle swarm optimization algorithm, swarm of particle includes 'n' number of particles, where position of each particle is responsible to indicate an individual solution in D-dimensional space [20]. Each & every particle in PSO is also associated with a velocity. All particles fly through a multidimensional search space where each particle is adjusting its position according to its own experience and that of neighbors. The particles change its condition according to the three principles: (a) inertia i.e. ω (b) it's most optimistic position i.e. P_{best} (c) swarm's most optimistic position i.e. G_{best} ; and converges to the most optimal position in the entire solution space by continuous change in the personal best and global best position [21].

An individual particle is treated as a volume-less particle in D -dimensional search space. The i^{th} particle is represented as $X_i = (x_i^1, x_i^2, \dots, x_i^D)$. Positions and velocities of the initial swarm of particles are randomly generated using upper and lower bounds on these search variables values, LB and UB [3]. X & V are calculated using the following formulae's, where Δt is the time step required. At each generation, each particle is updated by P_{best} and G_{best} . The P_{best} for i^{th} particle is represented as $P_i = (p_i^1, p_i^2, \dots, p_i^D)$ and is calculated as, And G_{best} at time step t is calculated by,

During iterations, movement of the particles is influenced either by its local neighborhood or the global neighborhood. That is, the portion of the adjustment to the velocity is influenced by the individual's previous best position is considered as the *cognition* component, whereas the portion influenced by the best in the neighborhood is termed as the *social* component [3][16][17]. With the addition of the inertia factor ω , brought in for balancing the global and the local search ability of the particles in PSO, these equations are:

Velocity Update Equation:

Position Update Equation:

Schematic representation of updating the velocity and position of a particle for optimal solution is shown in the Figure4:

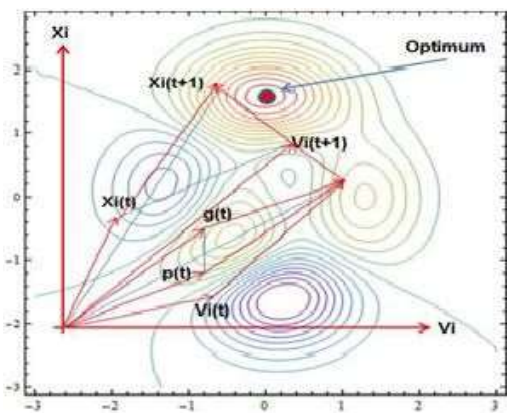


Fig 4: Schematic representation of updating the velocity and position of a particle for optimal solution[17]

Where $rand_1()$ and $rand_2()$ are two random numbers independently generated within the range $[0, 1]$ and c_1 and c_2 are two learning factors which control the influence of the social and cognitive components [3]. The basic steps of PSO algorithm are analyzed and given as below in Figure 5:

PSO Algorithm
1: initialize all particles according to (1) and (2)
2: for each particle do
3: calculate fitness for particle
4: if fitness better than previous P_{best} then
5: set fitness value as new P_{best} according to (3)
6: end if
7: end for
8: choose particle with the best P_{best} as G_{best} according to (4)
9: for each particle do
10: calculate particle velocity according to (5)
11: update the particle position according to (6)
12: end for
13: repeat
14: goto procedure 2
15: until maximum iteration or stop by user

Fig 5: Analyzed basic PSO algorithm [15][19]

V. EXPERIMENTAL ANALYSIS

Dataset Characteristics

In our literature survey, we came across Seven Data sets; those are namely, Vowel, Iris, Crude Oil, Contraceptive Method Choice (CMC), Glass, and Wine. Some authors have considered few or all of them in order to validate their proposed algorithms. These data sets are

of different dimensions such as high, low, and medium, which proved its importance whenever multidimensionality concept comes into consideration. The characteristics of these seven data sets can be well described in Table 2 and are easily available at link <http://archive.ics.uci.edu/ml/datasets.html>

Table 2. Data Sets Characteristics

Dataset	Total no. of instances	No. of Classes	No. of instances in each class	No. of Dimensions
Vowel	871	6	72,89,172, 151,207, 180	3
Iris	150	3	50,50,50	4
Crude oil	56	3	7,11,38	5
CMC	1473	3	629,334, 510	9
Cancer	683	2	444,239	9
Glass	214	6	70,17,76, 13,9,29	9
Wine	178	3	59, 71, 48	13

- (i) Vowel: This data set is a three dimensional data and is a collection of 871 Indian Telugu vowels sounds [1][7,8][10].
- (ii) Iris: Fisher Iris data set is a four dimensional data set, which is commonly, used in the pattern recognition problems. It has total 150 instances divided into 3 classes of iris flower of 50 instances each, where each class referring a type of iris plant (Iris setosa, Iris virginica and Iris versicolor) [1, 2, 3, 4][6, 7,8][10].
- (iii) Crude oil: This data set is taken from three zones of sandstone and has 56 instances with five attributes: vanadium, iron, beryllium, saturated hydrocarbons and aromatic hydrocarbons representing five dimensions[1][8].
- (iv) CMC: This data set is a nine dimensional data, a subset of 1987 National Indonesia contraceptive Prevalence survey. This is the sample of women who are married and were not pregnant or did not know if they were at the time of interview. The problem is to predict the choice of current contraceptive method of women on the basis of their demographic and economic characteristics[1][8].
- (v) Cancer: It is Wisconsin breast cancer data set that contains 683 objects with two types and nine attributes: clump cell size uniformity, thickness, marginal adhesion, cell shape uniformity, bare nuclei, single epithelial cell size, normal nucleoli, bland chromatin, and mitoses. The basic two types are malignant and benign [1][6, 7,8][10].
- (vi) Glass: The glass dataset contains the data collected from 6 different glasses: building windows float processed (70 objects), vehicle windows float processed (17 objects), building windows non-float processed (76 objects), containers (13 objects), tableware (9 objects), and headlamps (29 objects). Each data has nine features referring nine dimensions: sodium, refractive index, potassium, magnesium, calcium, silicon, aluminum, barium, and iron [1][4][10].
- (vii) Wine: This is the dataset that is obtained as the results of a chemical analysis of wines grown in the regions of Italy but derived from three different regions. The analysis determined the quantities of 13 constituents (inputs) found in each of the three types of wines [1, 2, 3,4][7][8].

Proposed Work

Data clustering with PSO algorithms have recently been shown to produce good results in a wide variety of real-world data. The variety of variations in PSO based clustering techniques are proposed in the literature achieving better results, although suffers from several limitations; more specifically when dealing with multidimensional data. This research propose a novel PSO based technique, which is a hybrid approach to the multidimensional data clustering problem by predicting the initial cluster center location and number of clusters and optimizing it with one of efficient PSO variant. A hybridized approach involves swarm intelligence inspired algorithm, (i.e.) BRAPSO [1] which is a variant of PSO with Subtractive Clustering algorithm [2]. The proposed approach is intended to design a system that will overcome the limitations associated with clustering where multidimensional data is of main concern. Hence, the proposed algorithm will resolve dead unit problem, stagnation, multiple cluster membership as well as premature convergence to local optima etc. The proposed work will use the data sets described in the previous subsection for experimental setup and will try to achieve better performance in terms of fitness value, error rate and convergence rate as compared to the existingsystems.

102

VI. CONCLUSION

The different aspects of Data Clustering, PSO and applicability of several PSO variants to data clustering are studied and analyzed, which helped in discovering the scope of research in PSO field when dealing with multidimensional data clustering. This idea motivated to propose a new PSO based clustering technique that can be a hybridization of Subtractive Clustering with BRAPSO

algorithm in order to achieve higher accuracy of clustering multidimensional data. Hence, the proposed technique is believed to defeat the key troubles linked with the existing PSO based clustering techniques such as hybrid subtractive + PSO clustering algorithm, BRAPSO algorithm, and the algorithms studied in the literature. The proposed algorithm is projected to exercise on seven multidimensional data sets to accomplish improved clustering outcome showing astonishing convergence performance by decreasing the error rate. Having the curse of dimensionality, the time complexity of the algorithm may increase linearly; therefore, the future work is towards handling this issue along with improvement in performance of clustering.

REFERENCES

- [1] S. Rana, S. Jasola, and R. Kumar, "A boundary restricted adaptive particle swarm optimization for data clustering," *International Journal of Machine Learning & Cyber.* Springer, June 2012, pp.391-400.
- [2] Mariam El-Tarabily, Rehab Abdel-Kader, Mahmoud Marie, Gamal Abdel-Azeem, "A PSO-Based Subtractive Data Clustering Algorithm," *International Journal of Research in Computer Science eISSN 2249-8265 Volume 3 Issue 2 (2013) pp.1-9.*
- [3] Sandeep Rana, Sanjay Jasola, and Rajesh Kumar, "A hybrid sequential approach for data clustering using K- Means and particle swarm optimization algorithm," *International Journal of Engineering, Science and Technology Vol. 2, No. 6, 2010, pp.167-176.*
- [4] Mehdi Neshat, Shima Farshchian Yazdi, Daneyal Yazdani and Mehdi Sargolzaei. "A New Cooperative Algorithm Based on PSO and K-Means for Data Clustering," *Journal of Computer Science 8 (2): 2012, ISSN 1549-3636. Pp.188-194.*
- [5] Serkan Kiranyaz, Turker Ince, Alper Yildirim, and Moncef Gabbouj, "Fractional Particle Swarm Optimization in Multidimensional Search Space," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 40, No. 2, April 2010, pp.298-319.*
- [6] Bara'a Ali Attea, "A fuzzy multi-objective particle swarm optimization for effective data clustering," *Springer-July 2010, pp.305-312.*
- [7] Suresh Chandra Satapathy & Anima Naik, "Efficient Clustering Of Dataset Based On Particle Swarm Optimization," *International Journal of Computer Science Engineering and Information Technology Research (IJCSSEITR) ISSN 2249-6831 Vol.3, Issue 1, Mar 2013, pp.39-48.*
- [8] Li-Yeh Chuang, Yu-Da Lin, and Cheng-Hong Yang, "An Improved Particle Swarm Optimization for Data Clustering," *IMECS Vol.1, 14-16 March 2012, pp.1-6.*
- [9] Xiaohui Cui, Thomas E. Potok, Paul Palathingal, "Document Clustering using Particle Swarm Optimization," *IEEE 2005, pp.185-191.*
- [10] R.Karthi, S.Arumugam, and K. Rameshkumar, "Comparative evaluation of Particle Swarm Optimization Algorithms for Data Clustering using real world data sets," *IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.1, January 2008, pp.203-212.*
- [11] Xiao-Feng Xie, Wen-Jun Zhang, and Zhi-Lian Yang, "Adaptive Particle Swarm Optimization on Individual Level," *IEEE, International Conference on Signal Processing (ICSP), Beijing, China, 2002, pp.1215-1218.*
- [12] Jianchao Fan,, Jun Wang, and Min Han, "Cooperative Coevolution for Large-scale Optimization Based on Kernel Fuzzy Clustering and Variable Trust Region Methods," *IEEE Transactions on TFS-2013-0157, pp. 1- 12.*
- [13] K. Premalatha and A.M. Natarajan, "Hybrid PSO and GA for Global Maximization," *ICSRS, Int. J. Open Problems Compt. Math., Vol. 2, No. 4, December 2009, pp.597-608.*
- [14] Chetna Sethi and Garima Mishra, "A Linear PCA based hybrid K-Means PSO algorithm for clustering large dataset," *International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013, pp.1559-1566.*
- [15] Ahmed A. A. Esmiin, Rodrigo A. Coelho and Stan Matwin, "A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data," *Springer, Feb 2013, pp.-1-23.*
- [16] Sandeep Rana, Sanjay Jasola, and Rajesh Kumar. 2010. A review on particle swarm optimization algorithms and their applications to data clustering. *Springer. (24 Nov 2010),211-222.*
- [17] Sunita Sarkar, Arindam Roy, and Bipul Shyam Purkayastha. 2013. Application of Particle Swarm Optimization in Data Clustering: A Survey. *International Journal of Computer Applications (0975 – 8887) Volume 65– No.25. (March 2013),38-46.*
- [18] Amreen Khan, Prof. Dr. N.G.Bawane, and Prof. Sonali Bodkhe. 2010. An Analysis of Particle Swarm Optimization with Data Clustering-Technique for Optimization in Data Mining. *(IJCSSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, (2010),1363-1366.*
- [19] Qinghai Bai. 2010. Analysis of Particle Swarm Optimization Algorithm. *Computer and Information Science, www.ccsenet.org/cis, Vol. 3, No. 1. (Feb 2010), 1-5.*
- [20] Ioan Cristian Trelea. 2003. The particle swarm optimization algorithm: Convergence analysis and parameter selection. *ELSEVIER. www.ComputerScienceWeb.com. Information Processing Letters 85. (2003),317–325.*
- [21] David Martens, Bart Baesens , Tom Fawcett. 2011. Editorial survey: swarm intelligence for data mining. *Springer, International Journal on Machine Learning. (2011),1–42.*
- [22] Aastha Joshi, and Rajneet Kaur. 2013. A Review: Comparative Study of Various Clustering Techniques in Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3. (March 2013),55-57.*
- [23] Khaled Hammouda. A Comparative Study of Data Clustering Techniques. www.pami.uwaterloo.ca/pub/-_hammouda/sde625-paper.pdf, 1-21.

- [24] Jiawei Han and Micheline Kamber. 2006. Data Mining Concepts and Techniques. Published by Morgan Kauffman, 2nd Ed, (2006).
- [25] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition. Springer Adaptation, Learning, and Optimization Vol.15.
- [26] Crina Grosan, Ajith Abraham, and Monica Chis. Swarm Intelligence in Data Mining. Springerlink,1-20.
- [27] Satyobroto Talukder. 2011. Mathematical Modelling and Applications of Particle Swarm Optimization. (Feb 2011).
- [28] Stuti Karol, and Veenu Mangat. 2013. Evaluation of text document clustering approach based on PSO. Cent. Eur. J. Comp. Sci., 3(2). (2013),69-90.