International Journal of Advanced Engineering Research and Science (IJAERS) *[Vol-3, Issue-11, Nov- 2016]*
https://dx.doi.org/10.22161/ijaers/3.11.20 *ISSN: 2349-6495(P) | 2456-1908(O)*

Performance Analysis of Genetic Algorithm with PSO for Data Clustering

G.Malini Devi[1], M.Lakshmi Prasanna[2], Dr.M.Seetha[3]

[1]Asst.Professor, Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad,
[2]M. Tech Student, Department of CSE, G. Narayanamma Institute of Technology and Science Hyderabad,
[3]Professor and HOD, Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad,

Abstract*—Data clustering is widely used in several areas like machine learning, data mining, pattern recognition, image processing and bioinformatics. Clustering is the process of partitioning or grouping of a given set of data into disjoint cluster. Basically there are two types of clustering approaches, one is hierarchical and the other is partitioned. K-means clustering is one of the partitioned types and it suffers from the fact that that it may not be easy to clearly identify the initial K elements. To overcome the problems in K-means Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) techniques came into existence. A Genetic Algorithm (GA) is one of hierarchical approach and can be noted as an optimization technique whose algorithm is based on the mechanics of natural selection and genetics. Particle Swarm Optimization (PSO) is also one of the hierarchical search methods whose mechanics are inspired by the swarming. The PSO algorithm is simple and can be developed in a few lines of code whereas GAs suffers from identifying a current solution but good at reaching a global region. Even though GA and PSO have their own set of strengths they have weaknesses too. So a hybrid approach (GA-PSO) which combines the advantages of GA and PSO are proposed to get a better performance. The hybrid method merges the standard velocity and modernizes rules of PSOs with the thoughts of selection, crossover and mutation from GAs. A comparative study is carried out by analyzing the results like fitness value and elapsed time of GA-PSO to the standard GA and PSO.*

Keywords*—*Clustering, GA-PSO, Genetic Algorithm, Particle Swarm Optimization.

I. INTRODUCTION

Clustering is a technique that is used to partition elements in a data set such that similar elements are grouped to same cluster while elements with different properties are grouped to different clusters as shown in “Fig 1” [7]. Clustering is a popular approach for automatically finding classes, concepts, or groups of patterns [4]. The reason behind clustering a set of data is to get a well structured data and expose this structure as a set of groups. It is used to perform efficient search of elements in a data set and is particularly effective in multi-dimensional data that may be otherwise difficult to organize in an effective manner. Such data is typically represented in the form of a floating-point number. We cannot use sorted arrays to search as they are multidimensional nature of data.

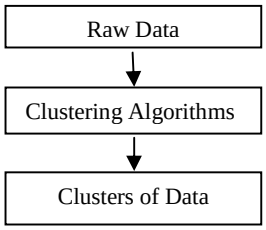

Fig.1: Clustering process.

Hash tables cannot be used because we may want to retrieve an item that is closest in properties to a specified item when the specified item does not exists in the data set. Clustering provides an elegant solution to this problem while providing a fast search capability for the same. Clustering widely used in areas like data mining [6, 9] image processing.

Most clustering algorithms belong to two groups: hierarchical clustering and partitioned clustering [3]. One of the partitioned clustering techniques in the literature is the K-means clustering method In this technique, clustering is based on the identification of K elements in the data set that can be used to create an initial representation of clusters[4]. These K elements form the cluster seeds. The remaining elements in the data set are then assigned to one of these clusters. Even though the method seems to be straightforward, it suffers from the fact that it may not be easy to clearly identify the initial K elements. To overcome

https://dx.doi.org/10.22161/ijaers/3.11.20 *ISSN: 2349-6495(P) | 2456-1908(O)*

the problem of partitioned clustering various heuristic algorithms have been proposed in the literature surveyed such as Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) [8].

GA is one of the hierarchical clustering algorithms and is inspired by biological system's improved fitness through evolution [7]. Genetic algorithms are based on three operations selection, crossover and mutation. It evolves a population of chromosomes representing potential problem solutions encoded into suitable data structures. Genes holds a set of values for the optimization variables[12]. To simulate the natural survival of the fittest process, best chromosomes exchange information (through crossover or mutation) to produce offspring chromosomes. The offspring solutions are then evaluated and used to evolve the population if they provide better solutions than weak population members. Usually, the process is continued for a large number of generations to obtain a best-fit (near optimum) solution.

The particle swarm optimization (PSO)[10] is a kind of optimization tool based on iteration, and the particle has not only global searching ability, but also memory ability, and it can be convergent directionally [1]. PSO is based on the behavior of a flock of migrating birds trying to reach an unknown destination. In PSO, each solution is a 'bird' in the flock and is referred to as a 'particle'. A particle is analogous to a chromosome (population member) in GAs. Physically, birds looks in a specific direction (towards their destination) and during their communication, they identify the bird that is in the best location. Accordingly, each bird speeds towards the best bird using a velocity that depends on its current position. PSO algorithm had basic three steps, namely, generating particles' positions and velocities, velocity update, and finally, position update. Each bird investigates the search space from its new local position, and the process repeats until the flock reaches a desired destination.

The proposed GA-PSO algorithm combines the features of both GA and PSO. It takes both the stability of the genetic algorithm and the local searching capability of Particle Swarm Optimization. The result proves that this method outperformed the GA and pure PSO in clustering efficiency.

II. LITERATURE SURVEY

2.1 Genetic Algorithm:

Current knowledge and many successful experiments suggest that the application of GAs is not limited to easy-to-optimize unimodal functions this work was proposed by Emmanuel Sarkodie Adabor *et al* (2012). The method Asymmetric key Encryption using Genetic Algorithm proposed by Poornima G.Naik ,Girish R. Naik *et al* (2013), describes an attempt to exploit the randomness involved in crossover and mutation processes for generating an asymmetric key pair for encryption and decryption of message. Tung-Kuan Liu, Yeh-Peng Chen and Jyh-Horng *et al* (2014) Chou says that Over time, the traditional single-objective job shop scheduling method has grown increasingly incapable of meeting the requirements of contemporary business models.

2.2 Particle Swarm Optimization:

Chuang *et al* (2012), suggest fresh particle swarm optimization (CPSO) algorithms that discover the best SNP arrangement for cancer connection studies containing seven SNPs. Marinakis *et al* (2013), this introduce a fresh algorithmic environment inspired techniques that uses a hybridized Particle Swarm Optimization algorithm with a fresh neighborhood topology for effectively solving the Feature Selection Problem (FSP). Akhshabi *et al* (2014), propose a particle swarm optimization (PSO) based on Memetic Algorithm (MA) that hybridizes with a local look for technique for work out a no-wait flow scheduling difficulty.

2.3 GA-PSO:

Optimal location management in mobile computing with hybrid Genetic algorithm and particle swarm optimization was proposed by Lipo Wang and Guanglin Si *et al* (2012). Priya I. Borkar and Leena H. Patil *et al* (2013) present a model of hybrid Genetic Algorithm -Particle Swarm Optimization (HGAPSO) for Web Information Retrieval. Yue-Jiao Gong, Jing-Jing Li, Yicong Zhou *et al* (2015) proposed that social learning in particle swarm optimization (PSO) helps collective efficiency, whereas individual reproduction in genetic algorithm (GA) facilitates global effectiveness.

III. GENETIC ALGORITHM [GA]

Genetic Algorithms are based on the concepts of natural selection and natural evaluation techniques [1]. Through reproduction genetic algorithm (GA) represents the evolution and improvement of life, when each individual holds its own genetic information through which a new one with fitness to the environment and more surviving chances is build. It is an iterative process and the evolution usually starts from a population of randomly generated individuals. The fitness of every individual in the population is evaluated at each generation; the fitness depicts the value of the objective function in the optimization problem being solved. The individuals are selected from the current

https://dx.doi.org/10.22161/ijaers/3.11.20 *ISSN: 2349-6495(P) | 2456-1908(O)*

population which is having best fitness value, and a new generation is formed modifying each individual's genome.

The fitness value of each individual is computed by the following fitness function. The fitness value is the sum of the intra-cluster distances of all clusters. This sum of distance has a profound impact on the error rate.

Fitness =∑ | Xj - Zi | , i=1,…., K j=1,……,n

Where *K* and *n* are the numbers of clusters and data sets, respectively. *Zi* is the cluster center at point *i* and *Xj* is the cluster for data point *j*.

The new individuals are formed using three genetic operators selection, crossover, and mutation. In selection individuals are selected based on their fitness value to generate offspring. The crossover aim of this mechanism was swapping to yield better fitness. Mutation increases the diversity and additional modifications increase the population [2]. These new offsprings are then used in the next iteration of the algorithm. Termination of the algorithm is occurred when either a maximum number of generations have been produced, or a satisfactory fitness level has been reached for the population. The flow steps of genetic algorithm for finding a solution of a given problem may be summarized as follows.

01: Begin
02: t=0
03: Initialize population P (t)
04: Evaluate fitness of each particle
05: t=t+1
06: If termination criterion occurs go to step 11
07: Select P (t) from P (t-1)
08: Crossover P (t)
09: Mutation P (t)
10: Go to step 4
11: Best output
12 Next generation until stopping criterion
13: End

The Genetic Algorithm consumes more CPU time. Here the CPU time is total time required to optimize a complete dataset by undergoing all the three operations like selection, crossover and mutation. As the algorithm is containing many operations to be done it requires more.

IV. PARTICLE SWARM OPTIMIZATION [PSO]

PSO was originally designed and introduced by Eberhart and Kennedy [6]. The PSO is a population search algorithm where each individual, called particle, within the swarm is represented by a vector in a multidimensional search space. A velocity vector is assigned to each particle to determine the next movement of the particle. Each particle updates its velocity based on the current velocity, best personal position it has explored so far and the global best position explored by the swarm [2]. The fitness function used to in this technique is as follows.

Fitness =∑ | Xj - Zi | , i=1,…., K j=1,……,n

Where *K* and *n* are the numbers of clusters and data sets, respectively. *Zi* is the cluster center at point *i* and *Xj* is the cluster for data point *j*.

Each particle is updated by following two "best" values at every generation. The first one is the best solution that has been achieved so far. This value is called *pbest*. Another "best" value which is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population. This value is a global best solution and called *gbest* [5]. After finding the two best values (*pbest* and *gbest*), the particle updates its velocity and position using the equations (1) and (2):

v(k+1) = *w* v(k) + c1 r1 (pbest(k) – pr(k)) + c2 r2 (gbest(k)- pr (k))….. (1)

pr(k+1) = pr(k) + v(k+1)…. (2)

where v(k) is the particle velocity; pr(k) is the current particle (solution) at the kth generation; r1 and r2 are two independent random numbers c1 and c2 are constants called acceleration coefficients; c1 controls the attitude of the particle of searching around its best location and c2 controls the influence of the swarm on the particle's behavior, and *w* is a constant known as inertia factor. Generally, the procedure for this algorithm is summarized as follows:

01: Begin
02: Initialize particles
03: While (number of iterations, or the stopping criterion is not met)
04: Evaluate fitness of each particle
05: For *n* = 1 to number of particles
06: Find *pbest*
07: Find *gbest*
08: For *d* = 1 to number of dimension of particle
09: Update the velocity and position of particles by equations (1) and (2)
10: Next *d*
11: Next *n*
12 Next generation until stopping criterion
13: End

The time taken to complete all the operations with a given number of iteration is known as CPU time. The PSO technique can execute a dataset with a satisfactory fitness value in less CPU time as the algorithm is simple and

https://dx.doi.org/10.22161/ijaers/3.11.20 *ISSN: 2349-6495(P) | 2456-1908(O)*

contains a very less operations like only updating the velocities current positional values.

V. HYBRID GENETIC ALGOTITHM WITH PSO [GA-PSO]

PSO often locates nearly optimal solutions at a fast convergence speed, but fails to adjust its velocity step size to continue optimization in the binary search space, which leads to premature convergence. In contrast, research has shown that genetic algorithms (GA) can adjust its mutation step size dynamically in order to better reflect the granularity of the local search area. However, GA suffers from a slow convergence speed. Although GAs have been successfully applied to a wide spectrum of problems, using GAs for large-scale optimization could be very expensive due to its requirement of a large number of function evaluations for convergence [11]. Therefore, hybrid GA-PSO has been proposed to overcome those problems and combine advantages of PSO and GA [13]. The basis behind this is that such a hybrid approach is expected to have merits of PSO with those of GA. One advantage of PSO over GA is its algorithmic simplicity. The idea behind GA is due to its genetic operator's crossover and mutation. The idea of combining GA and PSO is not new [5]. By applying crossover operation, information can be swapped between two particles to have the ability to fly to the new search area. Therefore, in our proposed hybrid GA-PSO, the crossover operation is also included, which can improve the diversity of individuals. Generally, the procedure for this algorithm is summarized as follows:

01: Begin
02: Initialize particles
03: While (number of iterations, or the stopping criterion is not met)
04: Evaluate fitness of each particle
05: For n = 1 to number of particles
06: Find *pbest*
07: Find *gbest*
08: Apply crossover and mutation operations
09: For d = 1 to number of dimension of particle
10: Update the velocity and position of particles by equations (1) and (2)
11: Next d
12: Next n
13 Next generation until stopping criterion
14: End Initialization

This hybrid approach consumes more CPU time because it need to perform all the operations of GA including the operations of PSO. But the fitness values generated by GA-PSO are more satisfactory than the other two approaches.

VI. EXPERIMENTAL RESULTS

In this section the three different optimization techniques are used to cluster ten different data sets and the results of GA-PSO are compared with standard GA and PSO algorithms in term of elapsed time and optimal fitness values. For the comparison purpose the stopping criteria that is number of maximum generations is taken same for all the three algorithms. Each algorithm will run 100 times.

Table.1: Results comparisons with Optimal Fitness Value

Datasets	**GA**	**PSO**	**GA-PSO**
Breast Cancer	1.90E+02	-154.3372	-185.9505
Concrete	1.49E+04	-122.7247	-147.2694
HayesRoth	8.233	-185.9505	-186.7039
HeartDisease	3.13E+04	-186.7309	-186.7039
Lung Cancer	5.34E+00	-122.8209	-144.3609
Seeds	-5.36E+00	-154.3372	-185.9003
Wine	1.87E+05	-28.885	-107.8089
Data	1.13E+04	-186.7309	-186.7309
Diabetic	1.33E+07	-184.4576	-185.6309
DataScience	2.01E+02	-185.1554	-185.1554

For the above Table 1 we can consider diabetic dataset as a best example because the GA-PSO can produce an optimized result than the other two and DataScience dataset is not performing well while using it with GA-PSO.

Table.2: Results comparisons with Time

Datasets	**GA**	**PSO**	**GA-PSO**
BreastCancer	32.560091	31.476014	48.668747
Concrete	31.573397	31.190026	31.452406
HayesRoth	31.254429	31.254429	32.744419
HeartDisease	49.111702	33.659704	51.12346
LungCancer	31.674727	31.453808	31.631851
Seeds	32.046989	31.816387	31.757602
Wine	32.956209	31.484516	32.359612
Data	169.634856	37.381394	140.483837
Diabetic	438.7682	224.48307	916.528801
DataScience	9797.553331	1592.40601	10876.302102

The above Table 2 stores the CPU time taken to complete the optimization for a dataset. Even though the GA-PSO gets the best optimized value it consumes more time.

VII. CONCLUSION

PSO works efficiently on large datasets by minimizing the time, utilizing the less parameter and gives the better

https://dx.doi.org/10.22161/ijaers/3.11.20 *ISSN: 2349-6495(P) | 2456-1908(O)*

performance than the GA by forming effective clusters. Proposed Hybrid (GA+PSO) methodology enhances the better performance results by incorporating the faster convergence and high computational speed than the individual comparison. The results show that the Hybrid of PSO and GA algorithms provides a performance that is significantly superior to that of other algorithm for these data sets. Genetic algorithm and particle swarm optimization are greatly related to their inherent parallel characteristics, both algorithms perform the function with a group of randomly created population, both have a fitness rate to calculate the population. PSO methodology is observed for document clustering limitation. It is found that the document clustering problem is successfully tackled with PSO methodology by optimizing for clustering process. A most useful advantage of the PSO is its capacity to cope with local optima by maintain, recombining and evaluation numerous candidate solutions concurrently. The Hybrid GA-PSO algorithm merges the capability of fast convergence of the PSO algorithm with the competence of ease to exploit preceding solution of GA for eliminating the early convergence.

REFERENCES

[1] Atul Garg, Dimple Juneja "A Comparison and Analysis of various extended Techniques of Query Optimization" *International Journal of Advancements in Technology, Vol. 3, No. 3, July 2012, ISSN 0976-4860, Page No.184-194.*

[2] Dr. Arvinder Kaur and Divya Bhatt "Hybrid Particle Swarm Optimization for Regression Testing" *International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 5, May 2011, ISSN 0975-3397, Page No.1815-1824.*

[3] E. Mehdizadeh, S. Sadi-nezhad and R. Tavakkoli-Moghaddam "Optimization of fuzzy clustering criteria by a Hybrid PSO And Fuzzy C-Means clustering algorithm" *Iranian Journal of Fuzzy Systems, Vol. 5, No. 3, 2008, Page No. 1-14.*

[4] K. Premalatha "A New Approach for Data Clustering Based on PSO with Local Search" *Computer and Information Science, vol 1, No 4, November 2008, Page No.139-145.*

[5] Lipo Wang and Guanglin Si "Optimal location management in mobile computing with hybrid Genetic algorithm and Particle Swarm Optimization (GA-PSO)" *IEEE Trans. Information Theory*, vol. 39, Page No. 1877-1886.

[6] Min Chen and Simone A. Ludwig "Fuzzy Clustering Using Automatic Particle Swarm Optimizations" *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) July 6-11, 2014, Page No 1545-1552.*

[7] Navpreet Rupal 1, Poonam Kataria "Comparative Analysis of Clustering & Enhancing Classification Using Bio- Inspired Approaches" *(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, 2014, ISSN: 0975-9646, Page No.6453-6457.*

[8] Oluleye Babatunde, Leisa Armstrong, Jinsong Leng, and Dean Diepeveen "Comparative Analysis of Genetic Algorithm and Particle Swam Optimization: An Application in Precision Agriculture" *Asian Journal of Computer and Information Systems, Vol. 03, Issue 01, February 2015, ISSN: 2321 – 5658, Page No.1-12.*

[9] P.Vivekanandan, R. Nedunchezhian **"**A Fast Genetic Algorithm for mining classification rules in large datasets" *International Journal on Soft Computing (IJSC), Vol.1, No.1, November 2010, Page No. 10-20.*

[10] Rahul Sharma "Comparative Analysis of Clustering by using Optimization Algorithms" *(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, 2014, ISSN: 0975-9646, Page No. 1076-1081.*

[11] Sapna Katiyar "A Comparative Study of Genetic Algorithm and the Particle Swarm Optimization" *AKGEC International Journal of Technology, Vol. 2, No. 2, Page No: 21-24.*

[12] Sameh Bennour, Amin Sallem, Mouna Kotti, Emna Gaddour, Mourad Fakhfakh, Mourad Loulou "Application of the PSO Technique to the Optimization of CMOS Operational Transconductance Amplifiers" *International Conference on Design & Technology of Integrated Systems in Nanoscale Era, 2014, Page No 1-5.*

[13] Sundararajan S and Dr.Karthikeyan " An Hybrid Technique for Data lustering Using Genetic Algorithm with Particle Swarm Optimization" *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 12, December 2014, ISSN: 0975-3397, Page No. 979-983.*