A Proportional Study of Issues in Big Dataclustering Algorithm with Hybrid Cs/Pso Algorithm

[1]A.C.Priya Ranjani ,[2]Dr. M.Sridhar
[1]Assistant Professor Department of Computer Science and Engineering,Vijaya Institute of Technology for Women Vijayawada, India.
[2]Associate Professor, Department of Computer Applications R.V.R & J.C College of Engineering Guntur, India.

ABSTRACT: - *Nowadays, it is very common in the framework of the cooperative system that theemerging need for distributed clustering algorithms is recognized to the massive size of databases. Though the distributed clustering algorithms enhance thedevelopment of integratingcalculations with communication and explain all the aspects of the distributed computing locations. Collaborative learning is the process in which several models are intentionallyproduced and shared to expose a specific computational intellectual problem. The proposed technique is able to find the optimal number of clusters spontaneously, in the situation of highly impossible trailing the number of clusters within the network. The proposedsystem is offered an efficient hybrid evolutionary algorithm based on mergingCuckoo Search (CS) and Particle Swarm Optimization (PSO) called as Hybrid CS-PSO algorithm. The supportiveframework is applicable to the cuckoo search algorithm to implement dimensional cooperation. The particle swarm optimization algorithm, viewed as a cooperative component, is embedded in the back of the cuckoo search algorithm. The experimental consequences determine that the proposed method is capable to attain a wellimproved big data clustering algorithm when compared with existing systems.*

Keywords: *- Cuckoo Search,Particle Swarm Optimization, Hybrid CS-PSO,big data clustering algorithm*.

I. INTRODUCTION

In the emerging world, the telecommunication networks and real-time control networks have the specified features of mobile nodes where the system can indulge into the outside world.Many relatingapproaches for clustering and organization are proposed and their confirmation methods too. Data mining means take outconcealed, formerstrange knowledge and rules with possible value to decision from mass data in database. Relationship rule mining is a main investigating area of data mining area, which is extensively used in repetition. With the growth of network technology and the enhancement of level of IT application, distributed database is commonly used. Clustering can be defined as the process of partitioning a set of patterns into separate and similar meaningful groups, called clusters. The emerging need for distributed clustering algorithms is renowned to the huge size of databases. Clustering is defined as the grouping a collection of entities into subgroups or clusters, such that those within one cluster are more closely associated to one another than objects allocated to different clusters, is a essential process of Data Mining. Especially, clustering is essential in informationacquirement. It is realistic in various fields comprising data mining, numerical data analysis, firmness and vector quantization. The mission of taking out information from large databases, in the custom of grouping rules, has fascinated extensive attention. The capability of numerous organizations to gather, stock and recover huge amounts of data has reduced the improvement of algorithms that can excerpt knowledge in the method of clustering rules. Distributed clustering algorithms hold this style of merging calculations with communication and sightsee all the aspects of the distributed computing environments. Consequently a distributed algorithm must yield under concern that the data may be fundamentally distributed to different roughly coupled sites connected through a network.

In the sections, the scalability of clustering techniques and the techniques for big data clustering much dynamic research has been committed.Big data clustering has been widely deliberated in many areas, together with statistics, machine learning, pattern recognition, and image processing. Different techniques has been introduced to overcome the problems happened in large database clustering, including initialization by clustering a model of the data and by means of an initial crude partitioning of the complete data set. Cuckoo search, a swarm intelligence-based nature-inspired algorithm, bases on the obligate brood parasitic behavior of some cuckoo species in combination with the Lévy flights behavior of some birds and fruit files. During iteration process, CS cooperatively uses Lévy flights random walk and biased/selective random walk to search new solutions. After each random walk, the greedy strategy selects a better solution from the current and new generated solutions according to their fitness. It has gained popularity due to its high efficiency. One of reasons for high efficiency may be that CS is a cooperative learner. Due to CS using multi-agents and sharing information among them, the individual cooperation is a basic form.

II. RELATED WORK

Zuo Chen et.al (2015) proposed A Self-Adaptive Wireless Sensor Network Coverage Method for Intrusion Tolerance Based on Particle Swarm Optimization and Cuckoo Search. In this paper, the combination of trust management model and heuristic optimization Particle Swarm Optimization and Cuckoo Search proposed a sensor network security coverage method based on trust management of intrusion tolerance. This method evaluates the trust value of the nodes through their behavior at first, and then adjusts the perception radius and decision-making radius. Finally, combine PSO and CS serial optimization in order to achieve the intrusion tolerance for efficient adaptive coverage. By comparing the simulation with a range WSN covering mechanism, this method has certain advantages over the performance of the algorithm, and in the case of the invasion can effectively protect the safety of the overlay network. The simulation results show the effectiveness of the algorithm.

Ching-Yi Chen et.al (2012) suggested Particle swarm optimization algorithm and its application to clustering analysis. Clustering analysis is applied generally to Pattern Recognition, Color Quantization and Image Classification. It can help the user to distinguish the structure of data and simplify the complexity of data from mass information. The user can understand the implied information behind extracting these data. In real case, the distribution of information can be any size and shape. A particle swarm optimization algorithm-based technique, called PSO-clustering, is proposed in this article. We adopt the particle swarm optimization to search the cluster center in the arbitrary data set automatically. PSO can search the best solution from the probability option of the Social-only model and Cognition-only model. This method is quite simple and valid and it can avoid the minimum local value. Finally, the effectiveness of the PSO-clustering is demonstrated on four artificial data sets.

M. ShamimHossain et.al (2016) proposed Big Data-Driven Service Composition Using Parallel Clustered Particle Swarm Optimization in Mobile Environment. The mobile environment is ambient and dynamic in nature, requiring more efficient techniques to deliver the required service composition promptly to users. Selecting the optimum required services in a minimal time from the numerous sets of dynamic services is a challenge. This work addresses the challenge as an optimization problem. An algorithm is developed by combining particle swarm optimization and k-means clustering. It runs in parallel using MapReduce in the Hadoop platform. By using parallel processing, the optimum service composition is obtained in significantly less time than alternative algorithms. This is essential for handling large amounts of heterogeneous data and services from various sources in the mobile environment.

Dang Cong Tran et.al (2013) provides a new approach based on enhanced PSO with neighborhood search for data clustering. The well-known K-means algorithm has been successfully applied to many practical clustering problems, but it has some drawbacks such as local optimal convergence and sensitivity to initial points. Particle swarm optimization algorithm (PSO) is one of the swarm intelligent algorithms; it is applied in solving global optimization problems. An integration of enhanced PSO and K-means algorithm is becoming one of the popular strategies for solving clustering problems. In this study, an approach based on PSO and K-means is presented (denoted EPSO), in which PSO is enhanced by neighborhood search strategies. By hybrid with enhanced PSO, it does not only help the algorithm escape from local optima but also overcomes the shortcoming of the slow convergence speed of the PSO algorithm.

III. PROPOSED SYSTEM

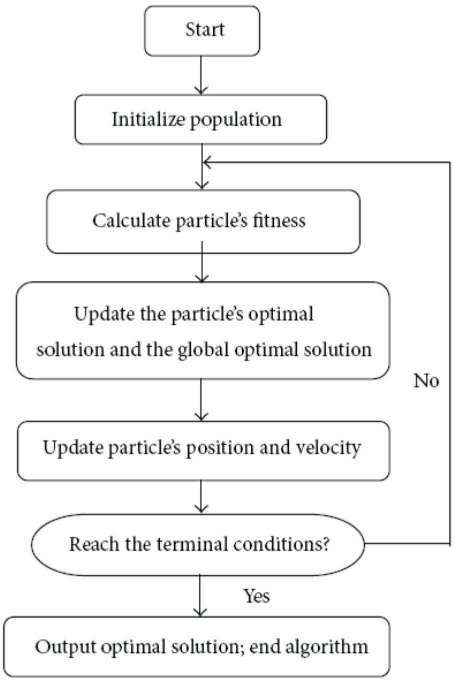
3.1 Particle Swarm Optimization

Particle Swarm Optimization is a population-based heuristic algorithm that is stimulated by the social actions of populations with cooperative properties. It emulates this kind ofcooperation and is commonly used in explaining mathematical optimization problems.As a whole, when PSO is engaged to solve an independent function f(x) with the dimensional search space, each unit i has two vectors to sustain its state, specifically, a position vector Xi(xi1,…,xiD) and a velocity vector Vi(vi1,…,viD). In the course of iteration, each unitappraises its velocity by learning from its general best position and the best position of the whole swarm acquired so far, and then informs its position according to its new updated velocity. The update rules in the original PSO are given as follows.

$$vid = vid + c1r1d(pBestid - xid) + c2r2d(gBestd - xid)$$

$$xid = xid + vid$$

Flowchart of Particle Swarm Optimization



3.2 Cuckoo Search

Cuckoo Search is a novel and well-organized population based heuristic evolutionary algorithm for resolving optimization problems. It has the benefits of simple implementation and few control limitations. This procedure is based on oblige brood opportunisticperformance of particular cuckoo types combined with the Lévy flight performance of some birds and fruit flies. It has been realistic to resolve a wide range of real-world optimization. In thosecomplications, such as structural optimization problem, shop planning problem, non-convexfinancial dispatch problem, and short-term hydrothermal scheduling problem.

The following are the approximation rules during the examine process

- ❖ Every cuckoo lays one egg at a time and dumps its egg in a arbitrarilyselected nest
- ❖ The best nests with a high-quality egg will be agreed over to the next generation.
- ❖ (3)The number of available host nests is fixed. A host bird can discover an alien egg with a probability $pa \in [0, 1]$.

In this situation, the host bird can either throw the egg away or abandon the nest and build a completely new nest.

In general, when it is used to solve an objective function f(x) with the solution search space $[xj, \min, xj, \max]$j = 1, 2,…,D, a nest represents a candidate solution $X = (x1, \ldots, xD)$.In the initialization phase, CS initializes solutions that are randomly sampled from solution space by

$$xi, j, 0 = xi, j, \min + r \times (xi, j, \max - xi, j, \min), i = 1,2, \ldots, NP$$

where r represents a uniformly distributed random variable with the range [0, 1], andNP is the population size.After initialization, CS goes into iterative phase. At generation G (G > 0), LFRW is firstly employed to search new solutions around the best solution obtained so far, and can be formulated as follows.

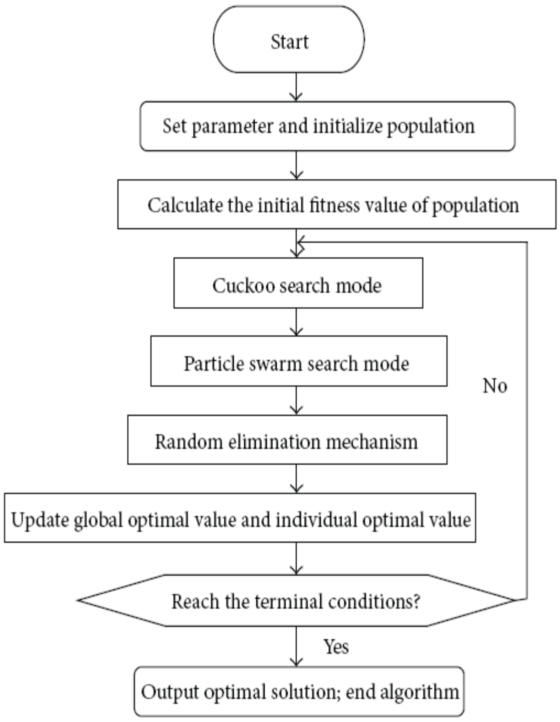
$$X_{i,G+1} = X_{i,G} + \alpha_0 \frac{\phi \times u}{|v|^{1/\beta}} (X_{i,G} - X_{best})$$

whereα0 is a scaling factor,
Xi,*G*means the *i*th solution at generation *G*,
*Xbest*represents the best solution obtained so far,
*u*and*v* are random numbers drawn from normal distribution with mean of 0 and standard deviation of 1.

3.3 Hybrid CS/PSO Algorithm

The Particle Swarm Optimization has benefits such as easy accepting, simple process, and prompt searching. On the other hand, in explaining a large complex problem, it becomes easily confined in local optimum. This fault must be astounded to spread the feasibility of algorithm. CS has advantages such as limited control restrictions and high efficiency, but it also has some defects, such as slow convergence speed and low accuracy. In CS, high randomness of the Lévyflight makes the search process quickly jump from one area to another area. Thus, the overall search talent of the algorithm is very strong. To recover the performance of CS, PSO is announced in the update process of CS. Thus, a CS-PSO hybrid algorithm is established. CS-PSO first uses Lévy flights in the search space to search, and then it uses the position of the PSO apprise mode to hasten the particles to the optimal solution convergence. Simultaneously, the arbitrary elimination mechanism of CS can effectivelyleak the local optima, thereby enlightening the performance of penetrating for the optimal solution. The procedure terms that are defined for the development of this function are Populace and Population Size, Fitness, Search Space Bound, Maximum Search Velocity, Minimum Search Velocity, PSO Search Mode, Cuckoo Search Mode and Discovery Probability.

Flowchart of Hybrid CS/PSO Algorithm



Step1: set the parameterand initialize the population.

Step 2: The initial fitness value of the population is intended by using the objective function, and the fitness value and location of the global optimum individual are specified.

Step 3: Cuckoo search mode is started. The fitness values of new and old individuals are compared. The enhancedoutcome is chosen as a new-generation distinct.

Step 4: PSO search mode is started. The position and velocity of the individual are rationalized, and then a new individual is created.

Step 5: In a contrast between the fitness values of the past and novel nests, the superior one will be preferred as a new generation of entities in the population.

Step 6: The universal and separate optimal values are rationalized. The optimal positions of all the individuals and whole populations are updated.

Step 7: If the end condition of the algorithm is satisfactory, then the optimal position of the nest is outputted, and the algorithm is terminated

IV. PERFORMANCE EVALUATION

In this proposed system, four test functions are selected to test the performance of the algorithm, and the results are associated with those of PSO and CS to verify the performance of the algorithm.The test functions $f1(x)$, $f2(x)$, $f3(x)$, and $f4(x)$, the global optimal solution found by PSO and CSO is not ideal with the current population size and number of iterations. To attain better results, the size of the greater population needs to be usual, and more iteration is required. The universal optimal solution create by CS/PSO is extremely close to the hypothetical optimumof the test function. By the average capability value, average optimal fitness value, and average standard deviation fitness value taken as the evaluating indicators, this paper tests PSOCS performance with changes in iteration times. The below graph demonstrates the optimal fitness value of the specified functions.

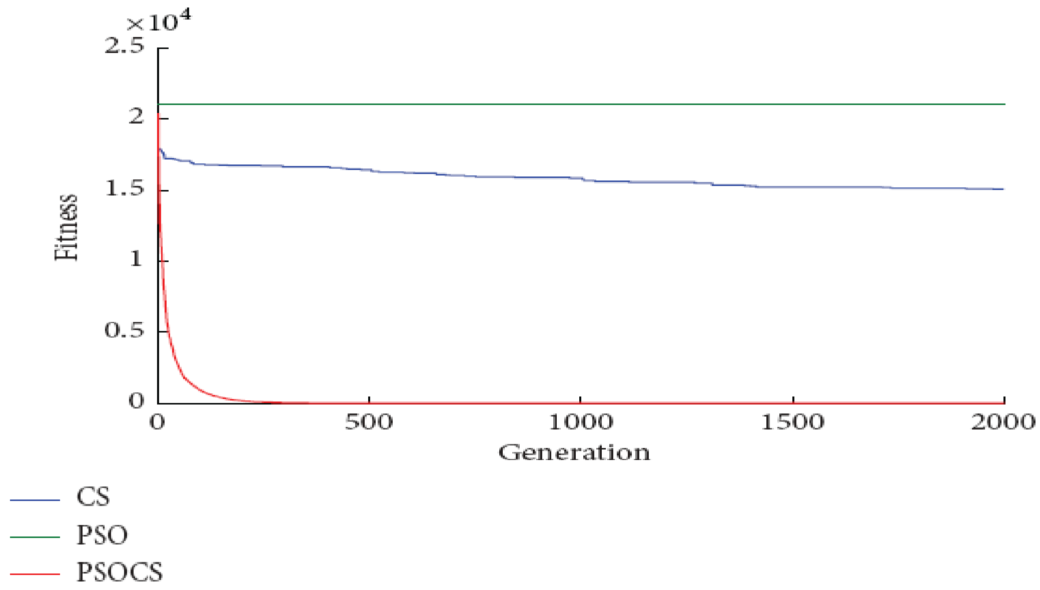


Fig.1: Change trend of function f1(x) average optimal fitness value

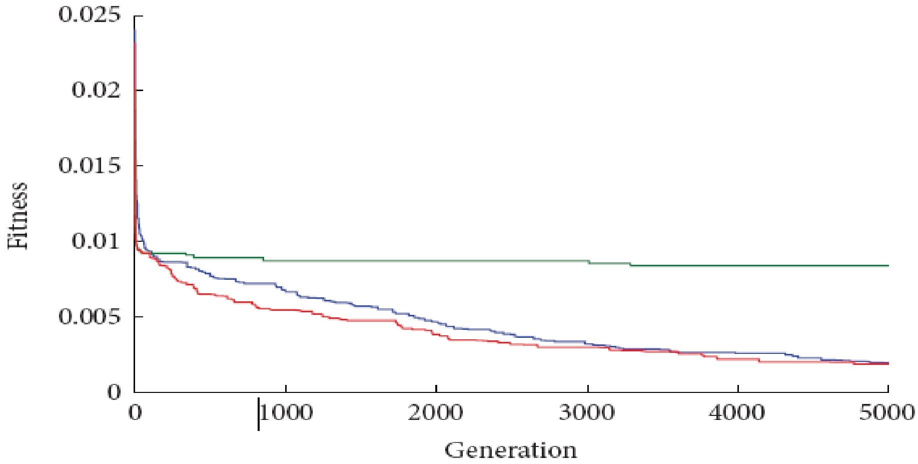


Fig.1: Change trend of function f2(x) average optimal fitness value

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a hybrid CS-PSO algorithm based on the PSO and CS algorithms. This mutual system combines the PSO algorithm's strong capabilityconcerningunion rate and the CS algorithm's strong capability in overall search. Consequently, it has a trade-off between eludingimpulsive convergence and discovering the whole search space. Compared with PSO and CS, hybrid CS-PSO has the benefits of fast convergence speed, strong penetratingcapability, and the ability to resolve the problem of multidimensional uninterrupted space optimization by using test tasks. Moreover, hybridization with other general algorithms such as PSO will also be possiblyproductive. More significantly, as for most meta heuristic algorithms, scientific analysis of the algorithm configurations is highly desired. Our investigationalassessmentestablished that the anticipated algorithm associatespositively to one existing algorithm on two multi-dimensional dataset. Investigational results exposed that the performance of this clustering algorithm is high, effective, and flexible.

REFERENCES

[1] JinchaoJi , Wei Pang, Chunguang Zhou, Xiao Han, Zhe Wang, "A fuzzy k-prototype clustering algorithm for mixed numeric and categoricaldata", journal of Knowledge-Based Systems, vol. 30, pp. 129-135, 2012.

[2] HesamIzakian, Ajith Abraham, Vaclav Snasel, "Fuzzy Clustering Using Hybrid Fuzzy c-means and Fuzzy Particle Swarm Optimization",World Congress on Nature and

Biologically Inspired Computing (NaBIC 2009), India, IEEE Press, pp. 1690-1694, 2009.
[3] GenlinJi and Xiaohan Ling, "Ensemble Learning Based Distributed Clustering", Emerging Technology in Knowledge Discovery and DataMining, Vol. 4819, pp 312-321, 2007.
[4] Noman, N, and Iba, H. (2008) 'Accelerating differential evolution using an adaptive local search',IEEE Transaction on Evolutionary Computation, Vol. 12, No. 1, pp.107–125.
[5] Wang, F., He, X.S., Luo, L.G. and Wang, Y. (2011a) 'Hybrid optimization algorithm of PSO and cuckoo search', Proc. of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), IEEE Press, pp.1172–1175.
[6] Wang, W.J. and Wang, H. (2014) 'An improved diversity-guided particle swarm optimization for numerical optimisation', Int. J. of Computing Science and Mathematics, Vol. 5, No. 1, pp.16–26.
[7] Kennedy, J. and Eberhart, R.C. (1995) 'Particle swarm optimization', Proc. of the IEEEInternational Conference on Neural Networks, IEEE Press, pp.1942–1948.
[8] Babukartik, R.G. and Dhavachelvan, P. (2012) 'Hybrid algorithm using the advantage of ACO andcuckoo search for job scheduling', International Journal of Information TechnologyConvergence and Services, Vol. 2, No. 4, pp.25–34.
[9] X.-S. Yang and S. Deb, "Cuckoo search via L´evy flights," in Proceedings of the World Congress on Nature & Biologically Inspired Computing, pp. 210–214, IEEE, Coimbatore, India,December 2009.
[10] X.-T. Li and M.-H. Yin, "A hybrid cuckoo search via L´evy flights for the permutation flow shop scheduling problem," International Journal of Production Research, vol. 51, no. 16, pp. 4732–4754, 2013.
[11] E. Vilian, S. Mohanna, and S. Tavakoli, "Improved cuckoo search algorithm for global optimization," International Journal of Communications and Information Technology, vol. 1, no. 1, pp 31–44, 2011.
[12] Zheng, H.Q. and Zhou, Y.Q. (2013) 'A cooperative co-evolutionary cuckoo search algorithm for optimization problem', Journal of Applied Mathematics, Article ID 912056.
[13] S.-S. Ma, H. Zhang, T. Feng, and J. Xue, "Optimization of preventive maintenance period based on hybrid swarm intelligence," in Proceedings of the 6th International Conference onNatural Computation (ICNC '10), vol. 5, pp. 2656–2659, IEEE, Yantai, China, August 2010.
[14] Yang, X.S.; Deb, S. Cuckoo Search: Recent Advances and Applications, Neural Comput. Appl.2014, 24, 169–174.
[15] X.-S. Yang and S. Deb, "Cuckoo search via L´evy flights," in Proceedings of the World Congress on Nature & Biologically Inspired Computing, pp. 210–214, IEEE, Coimbatore, India, December 2009