International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 9 (2017), pp. 2259-2265

http://www.ripublication.com

A Survey of Parallel Social Spider Optimization Algorithm based on Swarm Intelligence for High Dimensional Datasets

B.Shanmugapriya[1] and S.Meera[2]
[1]Research Scholar, Department of Computer Science,
[2]Assistant Professor, Department of Computer Science,
[1-2]PSGR Krishnammal College for Women, Coimbatore,
Tamil Nadu, India.

Abstract

Big data is the slightly abstract phase which describes the relationship between the data size and data processing speed in the system. The many new information technologies the big data deliver dramatic cost reduction, substantial improvements in the required time to perform the computing task or new product and service offerings. The several complicated specific and engineering problems can be transformed in to optimization problems. Swarm intelligence is a new subfield of computational intelligence (CI) which studies the collective intelligence in a group of simple intelligence. In the swarm intelligence, useful information can be obtained from the competition and cooperation of individuals. In this paper discussed about some of the optimization algorithms based on swarm intelligence such as Ant Colony optimization (ACO), Particle Swarm Algorithm (PSO), Social Spider Optimization (SSO) Algorithm and Parallel Social Spider Optimization (P-SSO) Algorithm. These optimization techniques are based on their merits, demerits and metrics accuracy, sum of intra cluster distance, Recovery Error Etc.

Keywords: optimization, swarm intelligence, ACO, PSO, SSO, P-SSO.

1. INTRODUCTION

The big data (Khan, N., et.al.2014) has gained much attention from the academia and IT industry. Because the companies should not store all the archives for long period nor more efficiently manage the huge data sets. Indeed the classical techniques has limited storage capacity, rigid management tools and so expensive. The lack of

scalability, flexibility and the performance is required in the big data context. The big data is mainly categorized based on the three aspects such as the data are numerous, the data cannot be categorized in to regular relational databases and the data are generated, captured and processed very quickly. The several complicated specific and engineering problems can be transformed in to optimization problems. The iterative development (Oussous, A., et.al. 2017) of the big data technologies the more and more difficult optimization problem are emerging and which has the relatively more objectives and large number of variables. The several optimization algorithms used such as Particle Swarm optimization (PSO), Ant Colony Optimization (ACO), Social Spider Optimization (SSO) and Parallel Social Spider Optimization (P-SSO) Algorithm.

In this paper the several optimization techniques to effectively optimize the large scale data based on their advantages, disadvantages and metrics.

2. EXISTING RESEARCH METHODOLOGIES

Social Spider Optimization Algorithm (Cuevas, E., et.al.2013) proposed social spider optimization algorithm for handle the optimization tasks. In present techniques the individual specified to emulate the group of spiders which is used to interact with each other depends on biological laws of the cooperative colony. Then the present's techniques only consider the two search agents (spiders) such as males and females. Based on the genders the specified task is evaluated for the various evolutionary operators which are minimizing the cooperative behaviour. In order to enhance the proficiency and robustness of the present technique this technique compared with the well known evolutionary techniques.

Hybrid Random / Deterministic Parallel Algorithm (Daneshmand, A., et.al.2015) proposed the decomposition framework for the parallel optimization for the sum of the various possibly non convex function and non smooth convex. The aim of this present technique is the novel parallel, hybrid random/deterministic decomposition scheme for each iteration and minimizing the function of convex surrogate of the original non convex. In order to handle the large scale problems the variables are updated based on the mixed random and deterministic procedure which is used to focusing the advantages of both pure deterministic and random update based techniques.

New Heuristic Optimization Approach (Hatamlou, A. 2013) proposed the new heuristic optimization approach which is inspired by the black hole phenomenon. Initially the present techniques start with the initial population of the candidate solutions for optimizing the objective function. Consider the each iteration of the black hole algorithm the based candidate is chosen to be the black hole. Assumed the each case the new star solutions is generated randomly then placed in the search space and start with the new search. In order to evaluate the black hole algorithm it's

applied to solve the clustering algorithm. The result shows that the performances are improved better than the classical heuristic algorithms.

Hybrid Social Spider Optimization (Chandran, T. R., et.al.2017) proposed new novel swarm intelligence called Social Spider Optimization (SSO) for the document clustering. The present technique provides better accuracy compare to the traditional optimization techniques. The present technique solves the document clustering problem. The performance parameters are threshold probability and the random variables effect clustering and deliver the effect of distance measure functions such as Euclidian and Manhatton functions on clustering. The static structure only implemented in the present technique so need the dynamic structure for automatically the text documents are added or removed when the number of clusters is not known. The performance are tested in terms of sum of inter cluster distance, sum of intra cluster distance and average cosine similarity.

Accelerated Distributed Rate Control Method (Chen, S., et.al.2017) proposed the accelerated distributed rate control methods for reducing the recovery error for the big sensory data. The present technique assures the minimization of the error recovery error of the reconstructed data and converges to the optimal value with the lower latency. In order to enhance the acceleration of the algorithm the present technique is build with the accelerated significant methods for the dual decomposition. The result shows that the convergence rate is improved compare to traditional optimization methods. Additionally the present techniques apply to the networks of the various sizes without sacrificing the acceleration gain.

Multi Objective Evolutionary Algorithm (Lin, F., et.al.2017) proposed multi objective optimization algorithm based on Non Dominated sorting and Bidirectional Local Search (NSBLA). The present techniques taken the local beam search as the main and provide the non dominated solutions via continuous iterative search while the iteration termination conditions are satisfied. The present techniques deliver the new distribution maintaining strategy based on the sampling theory to combine with the non dominated sorting algorithm to chosen the new population into next iterations. The present techniques deliver the better convergence for the Pareto optimal front compare to classical optimization techniques.

Social Spider Optimization (Vera-Olivera, H., et.al.2014) presents the Social spider optimization algorithm to optimizing the cluster of data. The all the outputs are validated using the non parametric wilcoxon rank sum test. The present techniques are evaluated over the 5 dataset from the UCI repository such as Balance, Cancer-Int, Dermatology, Diabetes and Iris and the comparison made between the k-means and genetic algorithm. The result shows that the present techniques deliver the competitive results and outperforms better than the K-means algorithm and genetic algorithm (GA).

Distributed Parallel Particle Swarm Optimization (Cao, B., et.al.2017) proposed the distributed parallel particle swarm optimization for improve the optimization. The

present techniques involves in the area of multi objective large scale optimization, many objective optimization and distributed parallelism. Initially the research are depends on the multi objective large scale distributed parallel PSO algorithms (MaOLSOPs). Then integrating in to novel variable decomposition and the optimization strategies then tracking the MaOLSOPs. When the objective functions are large the number of variables are huge then the optimization process also extremely time consuming. The results show that the present techniques improve the running efficiency.

Firefly Optimization Algorithm (Raj, E. D., & Babu, L. D. 2015) proposed firefly optimization Algorithm to overcome the privacy issue. The present techniques are generated by utilizing the data from face book and which can be incorporated in the any social networking website. The number of connections in the social networking site is increased extensively by utilizing the strategy of the present techniques. The firefly optimization algorithm used to increase the connections and enhance the interactions between the connections.

Computational Intelligence Optimization Algorithm (Cuevas, E., et.al.2015) proposed the novel swarm algorithm named Social Spider Optimization (SSO) to handle the optimization tasks. The present algorithm based on the simulations of the cooperative behaviour of the social spiders. The present algorithm the individual group of the spiders which is used to interact of each other depends on the biological laws of the cooperative colony. This technique considers only two agents such as the males and females. Based on the gender the specified is conducted by the set of the evolutionary operators which is used to mimic different cooperative behaviour which are typically found in the colony. To illustrate the proficiency and robustness of the present technique it's compare with the classical optimization technique.

Accelerated PSO Swarm Search Optimization Algorithm (Fong, S., et.al.2016) proposed a novel light weight feature selection to handle the high dimensional data and streaming format of data feeds in the big data. The present techniques mainly build based on the accelerated particle swarm optimization algorithm (APSO) to enhance the analytical accuracy with the reasonable processing time. In this technique, the collection of big data with the required large degree of the dimensionality are utilized under test the result shows that the present techniques enhance the performance.

Multi Objective Evolutionary Algorithm (Wang, H., et.al.2017) proposed the hybrid multi objective FA evolutionary algorithm (HMOFA) to solve the high dimensional optimization problems, and control the parameters are automatically adjusted while the searching process. Then the cross over strategy is used for monitoring the population diversity. The result shows that the present techniques reach the better performance compare to the traditional optimization techniques.

Parallel Social Spider Clustering Algorithm (Shukla, U. P., & Nanda, S. J. 2016) presents the Social spider optimization algorithm to enhancement of the clustering

algorithm and it's not favour the premature convergence or damage the exploration exploitation balance. This technique needs large number of iterations to predict the required convergence. So the new parallel SSO proposed to making the sequential movement of the dominate male, non dominate male and female spiders to parallel and reducing the computational complexity of the present techniques. In order to enhance the present techniques the simulation carried out to find out the nine small dimensional and eight high dimensional datasets. The superior performance of the optimized social spider optimization algorithm (O-SSO) demonstrated in the small data sets through the 'average classification error percentage' for the various computation algorithms such as APSO, ABC, RGA and K-means clustering algorithm.

3. COMPARATIVE ANALYSIS

The above survey delivers the detailed descriptions of various optimization techniques in big data. The following table shows that comparison of various optimization techniques in big data with their merits, demerits and metrics.

Table 1: Comparison Table

Ref No	Methods Used	Merits	Demerits	Metrics
[3]	spider optimization algorithm	Minimized the optimization tasks	Convergence is guaranteed but time to convergence uncertain	Minimization Results of benchmark functions=2.94E+00
[4]	Hybrid Random/Deterministic Parallel Algorithm	Enhancing the Convergence	Theoretical analysis is complex	Nil
[5]	New Heuristic Optimization Approach	Solver the NP hard problem	Complicate Coding	Sum of intra cluster distance=0.59940
[6]	Hybrid Social Spider Optimization	Speed of convergence, Linear time complexity	Poor cluster descriptors, Non convex clusters of varying size and the density cannot be handles	Accuracy=97.0626 Cosine similarity=0.9835
[7]	Accelerated Distribution Optimization Methods	Increase the convergence rate	Cannot be used in dynamic applications	Recovery Error=0.9
[8]	Multi Objective Evolutionary Algorithm	Provide the better convergence	Does not scale well with problem dimension	Index of Optimal Solutions= 2.70981 E-04
[9]	Social Spider Optimization	Enhance performance	High complexity	Median for Iris Dataset=97,222

[10]	Distributed Parallel Particle Swarm Optimization	Improve the efficiency	It does not allow to update model in online	Not Applicable
[11]	Firefly optimization Algorithm	Increase linking connection	Computational time	Number of Connections=14200
[12]	Social Spider Optimization	Improve performance	Need more computational time	5.1E-08
[13]	Accelerated PSO Swarm Search Optimization Algorithm	Improve efficiency	Low accuracy	Accuracy=0.790
[14]	Hybrid Multi Objective Evolutionary Algorithm	Improve the performance	Specifying weights for preference is not easy	Mean Value=6.71E+00
[15]	Parallel Social Spider Optimization (P-SSO)	Reduced Computational Complexity	Less Investigation	Accuracy=76.89%

4. CONCLUSION AND FUTURE SCOPE

In this paper, a detailed survey and further analysis on various optimization techniques in big data were encountered. It is obvious that all the authors have developed various optimization techniques to obtain the better results than the previous techniques with the suitable modifications. Even the parallel social spider optimization technique requires more investigation. To further research focus will be based on the Parallel Social Spider Optimization (P-SSO) to improve the accuracy for the handle the high dimensional tasks.

REFERENCES

[1] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M. ..,& Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. The Scientific World Journal, 2014.

[2] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2017). Big Data Technologies: A Survey. Journal of King Saud University-Computer and Information Sciences.

[3] Cuevas, E., Cienfuegos, M., Zaldívar, D., & Pérez-Cisneros, M. (2013). A swarm optimization algorithm inspired in the behaviour of the social-spider. Expert Systems with Applications, 40(16), 6374-6384.

[4] Daneshmand, A., Facchinei, F., Kungurtsev, V., & Scutari, G. (2015). Hybrid random/deterministic parallel algorithms for convex and nonconvex big data optimization. IEEE Transactions on Signal Processing, 63(15), 3914-3929.

[5] Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. Information sciences, 222, 175-184.

[6] Chandran, T. R., Reddy, A. V., & Janet, B. (2017). Text Clustering Quality Improvement using a hybrid Social spider optimization. International Journal of Applied Engineering Research, 12(6), 995-1008.

[7] Chen, S., Wang, K., Zhao, C., Zhang, H., & Sun, Y. (2017). Accelerated Distributed Optimization Design for Reconstruction of Big Sensory Data. IEEE Internet of Things Journal.

[8] Lin, F., Zeng, J., Lin, S., Zeng, W., & Lv, H. (2017). Multi-Objective Evolutionary Algorithm Based On Non-Dominated Sorting and Bidirectional Local Search for Big Data. IEEE Transactions on Industrial Informatics.

[9] Vera-Olivera, H., Soncco-Alvarez, J. L., & Enciso-Rodas, L. Social Spider Algorithm Approach for Clustering.

[10] Cao, B., Zhao, J., Lv, Z., Liu, X., Yang, S., Kang, X., & Kang, K. (2017). Distributed Parallel Particle Swarm Optimization for Multi-Objective and Many-Objective Large-Scale Optimization. IEEE Access.

[11] Raj, E. D., & Babu, L. D. (2015). A firefly swarm approach for establishing new connections in social networks based on big data analytics. International Journal of Communication Networks and Distributed Systems, 15(2-3), 130-148.

[12] Cuevas, E., Cienfuegos, M., Rojas, R., & Padilla, A. (2015). A Computational Intelligence Optimization Algorithm Based on the Behaviour of the Social-Spider. In Computational Intelligence Applications in Modeling and Control (pp. 123-146). Springer International Publishing.

[13] Fong, S., Wong, R., & Vasilakos, A. V. (2016). Accelerated PSO swarm search feature selection for data stream mining big data. IEEE transactions on services computing, 9(1), 33-45.

[14] Wang, H., Wang, W., Cui, L., Sun, H., Zhao, J., Wang, Y., & Xue, Y. (2017). A Hybrid Multi-Objective Firefly Algorithm for Big Data Optimization. Applied Soft Computing.

[15] Shukla, U. P., & Nanda, S. J. (2016). Parallel social spider clustering algorithm for high dimensional datasets. Engineering Applications of Artificial Intelligence, 56, 75-90.