Artificial Bee Colony Algorithm for Optimizaion in Data Science

Combination of Artificial Intelligence and Data Science

Mr. Shaleen Shukla, Prarthana Fadia,
Assistant Professor, Assistant Professor
Information Technology, Gyanmanjari Institute of Technology, Bhavnagar, India

Abstract— Data science is all about performing various operations on various type of data. Big data is a large amount of data which is hard to handle by on hand systems. It requires new structures, algorithms and techniques. As data increases as per volume, dark data also will increase. Artificial Bee Colony algorithm is a part of Swarm Intelligence. It is based on how honey bees are working to find out their food sources. In Big Data there is distributed environment so required sources may be on different places. During process the data these data sources have to find out from different places and analyze a one system. This requires calculation which can help us to find out best option for our required data sources. ABC algorithm is used to overcome limitations of ant colony algorithm. In ant colony initialization will be repeat from starting point in case of failure. In bee colony optimization initialization happens only once. It is used to find out required data source based on parameters out of multiple data sources. Thus, artificial bee colony algorithm can be used to find out best data sources. We can store these derived data sources on cloud for further processing. Bee colony algorithm generally used in data mining and networking field. It can be used for Big Data for identifying data resources.

Index Terms— Ant colony optimization; Bee colony optimization; Distributed data sources

I. INTRODUCTION

Here study is about to combine Big data with artificial bee colony optimization. Artificial bee colony optimization is used in networking and data mining. In networking field, artificial bee colony optimization is used to find out best destination address via shortest path. Artificial bee colony optimization was developed in 2005 and its main characteristic is to find out best data source from given multiple data sources. It is a part of swarm intelligence and discovered to overcome limitations of an ant colony optimization. It works on probabilities based on fitness of sources/destinations. [1] Fitness is decided based on different parameters. Source satisfying maximum parameters is most useful and sufficient source. Probability is ratio of fitness of any source to total fitness of all sources within a given system. Source with highest probability is chosen for process first.

In Artificial Bee Colony (ABC) optimization behavior of honey bees is taken as a reference. There are main 4 types of bees. [1]

- Queen(it lives always in a hive)
- Onlooker(observe work of workers and scouts)
- Worker(find food sources initially)
- Scout(once food source of any worker is exhausted then it becomes a scout and try to find other)

Here is method that how they are finding food source. [1]

- **Worker:** worker bees are traveling around and try to find a food source. Once they got a food source, all bees informs each other by performing dance called “waggle dance.”
- **Scout:** once a food source of any bee is exhausted by other bees then it becomes a scout and go to find a new source. Thus we can say that scout is nothing but an experienced worker bee.
- **Onlooker:** onlooker observes all these activities of scouts and workers by staying nearer of hive. Initially routes to travel for all bees are also decided by onlooker. Workers give all information about food source to onlooker so onlooker will send rest all bees to that direction only.

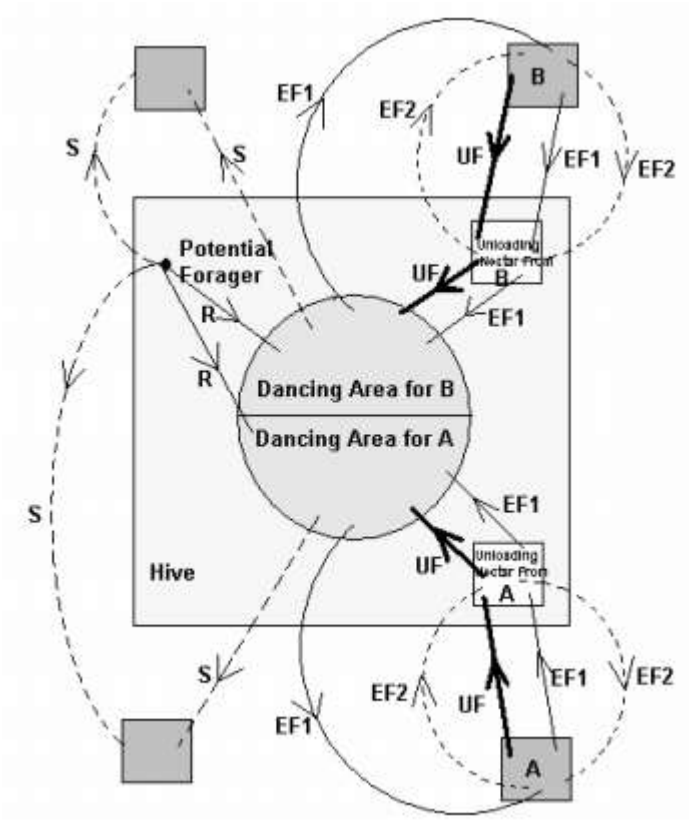


Fig. 1.1 The behavior of honey bee foraging for nectar [8]

II. LITERATURE REVIEW

Big data has very large amount of data and sometimes clusters or classes created from it also may be very large. In such situation if we want to apply ABC algorithm on it, we should create an environment in which our optimization process can work on large data sets. For this purpose we can chose to work on cloud platform or other platforms like Hadoop and Spark. Generally HDFS and spark can handle big data, but if we want to perform ABC optimization on any data set temporarily then we can use cloud storage as SaaS(software as a service), PaaS(platform as a service) and IaaS(infrastructure as a service). [3] In this way we can store clusters on cloud with some predefined fitness to perform operations to find out the best one.

ABC is well suited for general assignment problem, cluster analysis, constrained problem optimization, structural optimization, and advisory system. It has also been applied to software engineering for software testing and parameter estimation in software reliability growth models. Thus ABC is efficient applicable algorithm for rising technologies. [1]

Cloud computing is useful to process ABC optimization. In such cases we can load data sets through VMs. ABC helps in cloud computing in case of load balancing. ABC can identify only required best data source and load it n cloud to reduce load. Whenever we want to perform more number of processes, ABC also can help in VM migration. Memory utilization and processing speed can be manage and improve by using ABC optimization. [2]

Big data consists large quantity of data and still it increasing day by day. Operations perform on it also requires high capacity of processing, high level of memory utilization and resource migrations. In this way cloud computing is like ray of hope. Cloud storage provides facilities to store and process data virtually without any load on main processing system. Data we want to analyze for ABC optimization also can put on cloud environment temporarily. [3] In fact we can say that today Big Data and cloud computing increases in parallel ways and they may also become very important supports for each other.

Here is way how ABC used in wireless communication. It is suitable for wireless sensor network. Here, nodes in same communication station will be divided into clusters. Each cluster has its own worker bees (communicating nodes) and onlooker (cluster head). Cluster head can act as an interpreter with other cluster heads from different clusters for communication purpose. They can act same as honey bees to find out their destination nodes.[4]

Ant colony optimization is useful for smaller distance and smaller datasets only. First of all ants will be initialized. Then some out of multiple ants travelling through multiple paths as a worker to find sources and leave pheromones after themselves and other ants can follow them by sensing pheromones left by previous ants. Thus all ants can reach to required food source.[5]

III. RELATED WORK

Ant colony optimization is a part of swarm intelligence and it is generally used in wireless communication system and data mining where distance and data is short. It is based on how ants are working to find out their food sources. But there are some limitations in ant colony optimization. Like if one ant will found bad source then all other ants will be misguide by it. In other hand if found source is not sufficient hen re-initialization of ants is must. Thus it is more time consuming process so it can use only in case on less distance or less amount of data.[5]

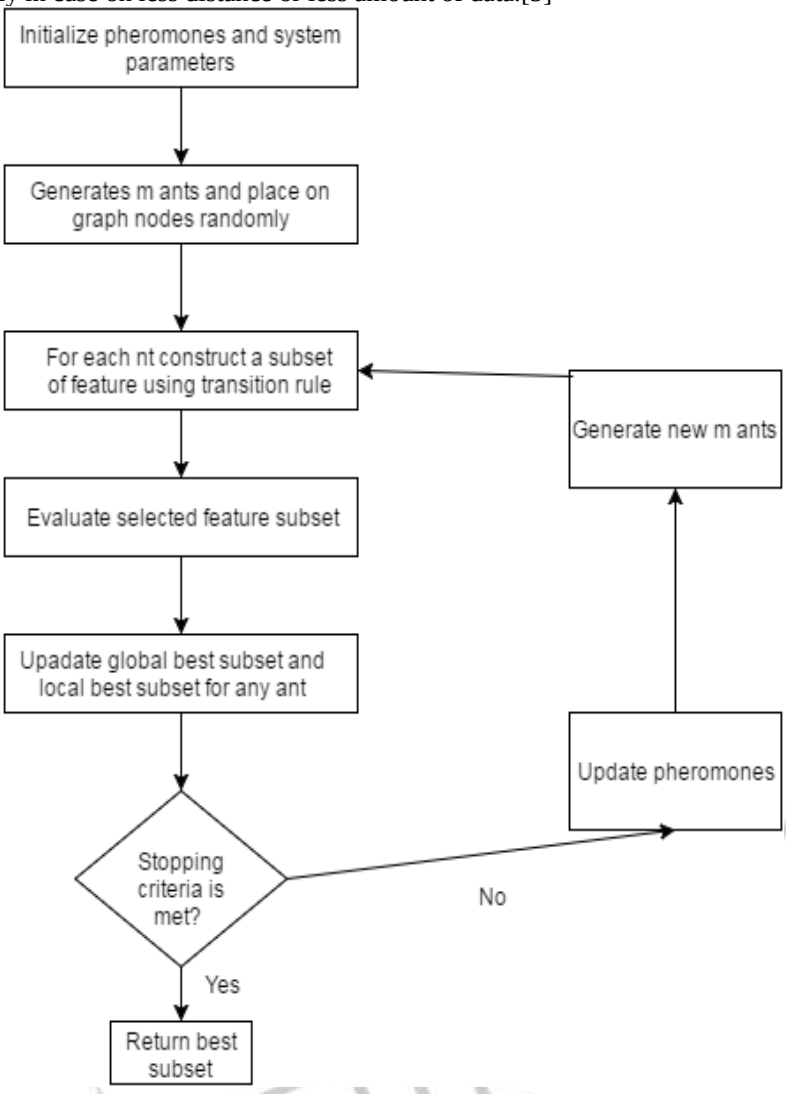


Figure 1. Flow chart of ant colony optimization

IV. PERFORMANCE EVALUATION

Bee colony optimization is a part of swarm intelligence. Before ABC introduced, ant colony optimization was popular to find out best source from given environment. Ant colony optimization is works based on ants' behavior to find out sources. [5] Characteristics of ant colony are..

- Suitable for smaller size of data
- Slow to deliver result
- Less accurate in case of multiple data sources
- Re-initialization is required if delivered data source is not sufficient.

Psudo code example for ant colony optimization [11]

```
Proceure ACO-Metaheuristic
        While(not-termination)
                generateSolutions()
                daemonActions()
                pheromoeUpdate()
        end while
end procedure
```

In ant colony optimization, initially out of all ants some ants starts travelling through different edges to find out sources. Each ant spreads pheromones so that rest all ants can follow that path. In this way each ant leaves pheromones to guide rest all ants. If any path stays ideal for some time without travelling of any ant, pheromones evaporate automatically. Thus, path with strongest pheromones is declared as a best path to destination. But in this case destination decides only based on how first worker ant found source. Ant finding a source in unaware from probability of other sources. Thus if once wrong source is found, rest all ants will be misguided to it without finding other options. This is a main disadvantage of ant colony optimization. In this way ant colony optimization is less accurate in case of larger number of data sources.

To overcome limitations of ant colony optimization, artificial bee colony algorithm is introduced. It is based on method of honey bees for finding sources. The employed bees share the information about their food sources with onlooker bees after all of them complete the search process. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount by Eq. known as roulette wheel selection method which provides better candidates to have a greater chance of being selected. [4]

Characteristics of artificial bee colony optimization are…

- Suitable for larger size of data
- Fast to deliver result as compare to ant colony optimization
- More accuracy also in case of multiple data sources
- Re-initialization isn't required if delivered data source is not sufficient
- It can easily accept new data source

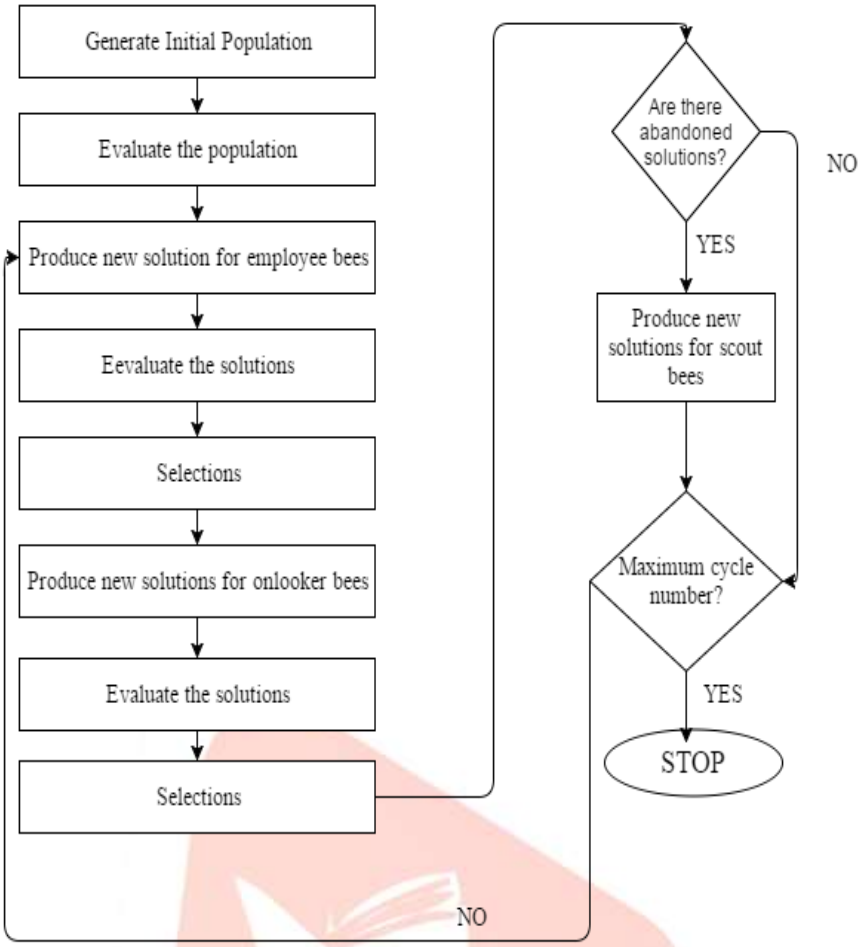
To choose data source based on fitness they should be clustered w.r.t. some parameters. Parameters used to decide fitness of any data source are different for different data source. It means based on different cased it is necessary to choose different parameters. Purpose of finding data source will be helpful to define parameters.

ALGORITHM [4]

1. Generate initial population Xi, I = 1…SN
2. Evaluate the population
3. Set cycle to 1
4. Repeat
5. FOR each employed bee
6. Produce new solutions vi by using
7. Calculate the fitness
8. Apply the greedy selection process
9. FOR each onlooker bee
10. Choose a solution xi depending on pi
11. Produce new solutions vi
12. Calculate the fitness
13. Apply the greedy selection process
14. If there is an abandoned solution then
15. Replace it with a new solution produced by a scout using
16. Memorize the best solution achieved so far
17. cycle = cycle ? 1
18. Until cycle = MCN

The employed bees share the information about their food sources with onlooker bees after all of them complete the search process. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount by equation known as roulette wheel selection method which provides better candidates to have a greater chance of being selected.

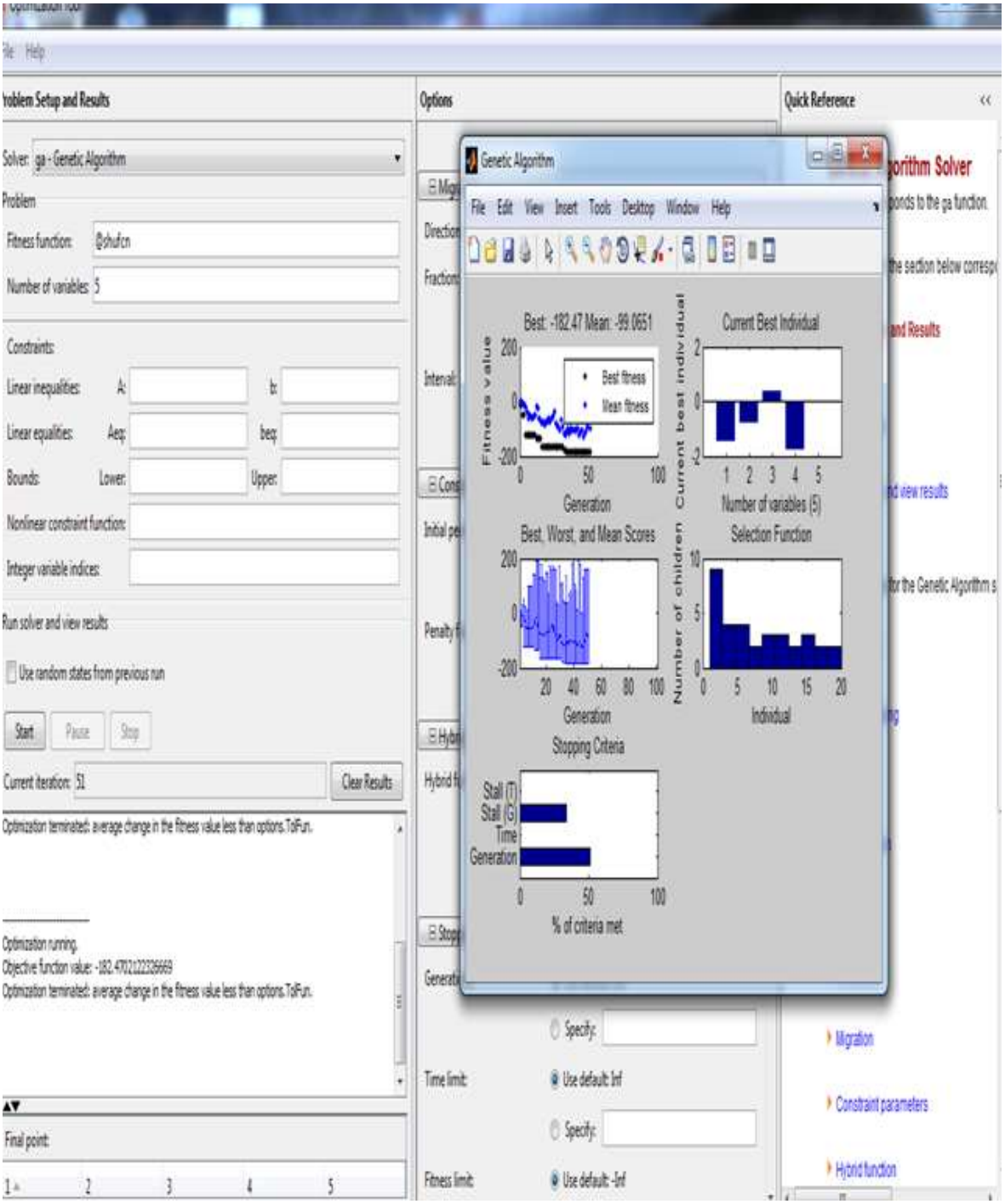
$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n}$$



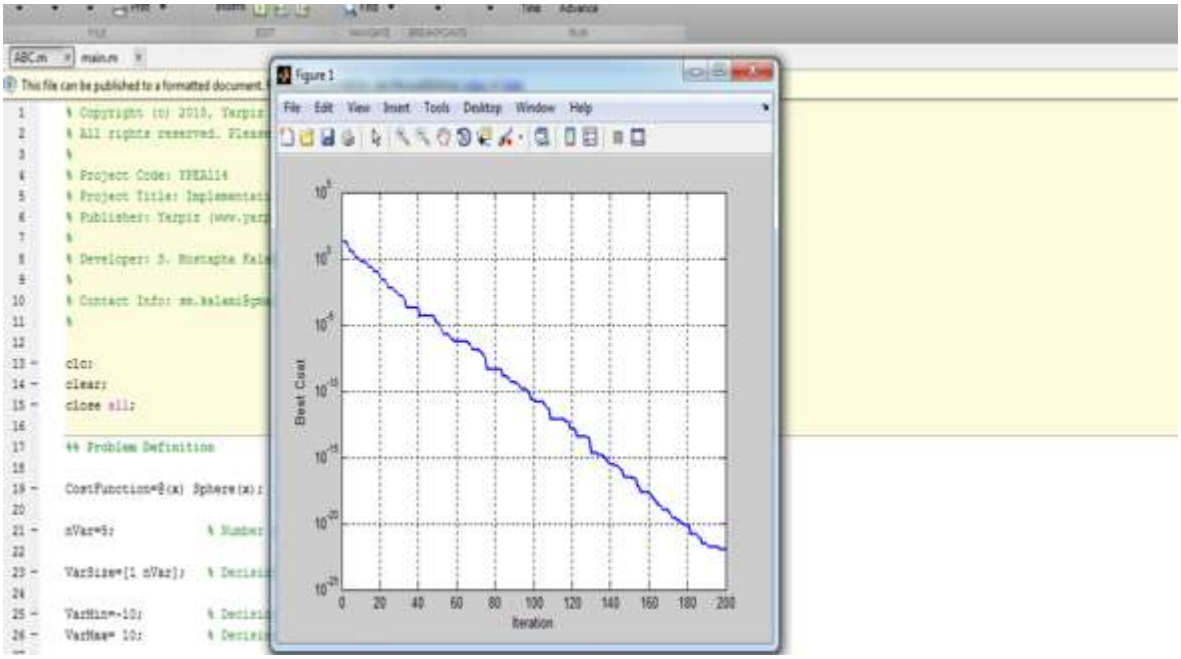
Figur 2. Flowchart of artificial bee colony optimization

V. METHODOLOGY

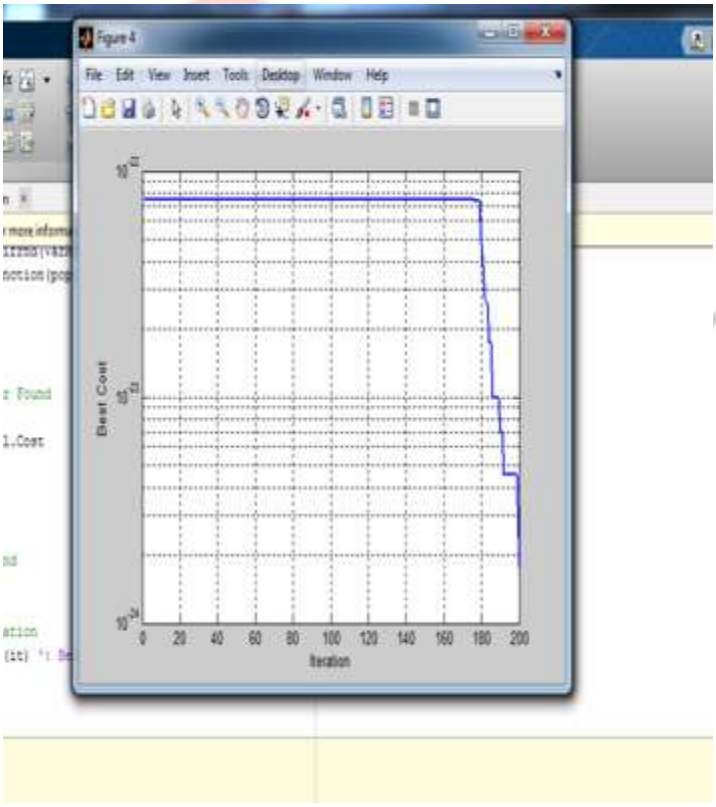
Here are some practical outputs derived in matlab. First of all here is an output in matlab for genetic algorithm with 5 variables for optimization.



Here is output of code for ABC developed in matlab for optimization purpose. Here time optimization is taken under consideration. Code highlighted in image below shows code for generating initial population. Here 5 variables are taken for testing purpose.



Output-1



Output-2

Here genetic algorithm took 11 seconds to complete process and give output for 5 variables. But ABC algorithm gave output for 5 variables in 4 seconds. Thus ABC algorithm is very fast for optimization techniques.

VI. Conclusion

Here, ABC optimization can help us to find out required data source from large amount of data sources, ABC also can help to arrange output of Big Data in particular pattern. If clusters, analyzed output of data sets, selected data sets will be on cloud storage or on different location, ABC optimization is easy way to determine them. Spark platform is useful to handle large data set in the form of clusters. ABC is good in optimization. ABC optimization will be different in different manners because fitness can be different in different cases due to different parameters.

VII. Acknowledgements

This research was supported by Staff of Parul University, Gyanmanjari Institute of Technology and Silver Oak College of Engineering and Technology.

REFERENCES

Researh Papers

[1] "Applications of Artificial Bee Colony Optimization Technique: Survey", Kuldeep Singh Kaswan, Sunita Chaudhari, Kapil Sharma, 978-9-3805-4416-8/15/$31.00 c 2015 IEEE

[2] "Enhanced Bee Colony Algorithm for Efficient Load Balancing and Scheduling in Cloud", K.R. Remesh Babu and Philip Samuel, Springer International Publishing Switzerland 2016

[3] "The rise of "big data" on cloud computing: Review and open research issues", Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar,Salimah Mokhtar , Abdullah Gani, Samee Ullah Khan, Information Systems 47 (2015) 98–115

[4] "Cluster based wireless sensor network routing using artificial bee colony algorithm Dervis Karaboga, Selcuk Okdem, Celal Ozturk, Springer Science+Business Media, LLC 2012

[5] "Optimization of the Running Speed of Ant Colony Algorithm with Address-based Hardware Method", ElnazShafighFard, Khalil Monfaredi, ISSN: 2180 - 1843 Vol. 7 No. 1 January - June 2015

[6] "Distributed Virtualization Manager for KVM Based Cluster", Mr. Uchit Gandhi, Mr. Mitul Modi, Ms. Mitali Raval, Mr. Paavan Maniar, Dr. Narendra Patel, Prof Kirti Sharma, Procedia Computer Science 79 (2016) 182 – 189, ScienceDirect

[7] "Data Model for Big Data in Cloud Environment", Imran Khan, S. K. Naqvi, Mansaf Alam, S. N. A Rizvi

[8]"Research of Resource Allocation in Cloud Computing Based on Improved Dual Bee Colony Algorithm", Wu Ju-Hua, International Journal of Grid Distribution Computing, Vol. 8, No.5, (2015), pp.117-126

[9]"SAACO: A Self Addictive Ant Colony Optimization in Cloud Computing", Weifeng Sun, Zhenxing Ji, Jianli Sun, Ning Zhang, Yan Hu, 2015 IEEE Fifth International Conference on Big Data and Cloud Computing

[10]"Unsupervised probabilistic feature selection using ant colony optimization", Behrouz Zamani Dadaneh*, Hossein Yeganeh Markid, Ali Zakerolhosseini, Expert Systems With Applications 53 (2016)

[11] "AN IDEA BASED ON HONEY BEE SWARM FOR NUMERICAL OPTIMIZATION ", Dervis KARABOGA, (TECHNICAL REPORT-TR06, OCTOBER, 2005)

[12] "Implementation of Artificial Bee Colony Algorithm", IAES International Journal of Artificial Intelligence (IJ-AI) VOl1 September 2012.

Websites

[13] https://en.wikipedia.org/wiki/Artificial_bee_colony_algorithm

[14] https://en.wikipedia.org/wiki/Swarm_intelligence

` [15] http://mf.erciyes.edu.tr/abc/index.htm

[16] https://en.wikipedia.org/wiki/Ant_colony_optimization_algorithms