International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307

An Efficient Survey of Managing Big Data Analytics using Swarm Intelligence

V. Ramesh
CSE Department, Assistant Professor, Sri Indu College of Engineering & Technology, Hyderabad, Telengana, India

ABSTRACT

Within the current scenario, the big data analytics is wide within the discussion. The characteristics of massive information are high volume, high variety, and high velocity. These characteristics create massive information analysis difficult. Challenges visages are high dimensionality, dynamically ever-changing information. Swarm intelligence has a capability to resolve dynamical, huge, and multi-objective issues. Here we focus on proving that various huge data analytics problem is solved using swarm intelligence and it is applied to Hadoop design. During this paper, we use a particle swarm optimization rule to form clusters of given dataset. Several huge data analytics issues are solved using swarm intelligence technique.

Keywords: Big Data Analytics, Swarm Intelligence (SI), Particle Swarm Optimization (PSO).

I. INTRODUCTION

Big data consists of very large or complex data sets that the data processing applications currently available are inadequate to deal with. The challenges faced include storage, transfer, sharing, querying, updating, search, analysis, data creation, and capture and information privacy.

Big data has some specific characters and are defined by 7V's of big data, which sometimes are called the spectrum of big data.

- ✓ Volume: This is the size of big data. It is very high and is difficult to manage this huge data in traditional database systems.
- ✓ Velocity: Velocity is the rate at which the data is being generated continuously from different sources.
- ✓ Variety: Big data is very heterogeneous in nature. This data can be structured, semi-structured and unstructured. Data from different sources is being generated that has different formats and are stored to process and analyze.
- ✓ Validity: As the volume of big data is very high, it is our major requirement to get accurate and relevant data for our analysis purpose.
- ✓ Veracity: not all the data stored in databases is useful. It has abnormality and redundancy in large amount. To make this data more informative and clean is a major challenge for our analysis purpose.
- ✓ Volatility: The most critical issue for big data is data storage. The validity of data and the duration of data to be stored in the databases are the issues related to volatility of big data. The volatility issues of the data need to be fixed for our analysis purpose.
- ✓ Viability: It refers to choose the appropriate factors and relevant features of the considered data set. This contributes maximum in predicting the result of analysis.

Swarm intelligence was first introduced by Beni. Swarm intelligence is defined as the collective behavior of decentralized, self-organized natural or artificial systems, which is based on a population of individuals where each individual represents a potential solution of a problem being optimized [one]. Swarm intelligence is sub field of artificial intelligence. The nature of swarms called as agents gave birth to the swarm intelligence technique. For example, the ability of bees and ants to orient themselves in the environment. Swarms can interact both locally as well as globally. Ant colonies, bacterial foraging, birds flocking are the swarm intelligence which exists in nature. Principles followed by all swarms are principle of proximity, principle of quality, principle of diversity, principle of stability, principle of adaptability.

- ✓ Principle of proximity: It states that during the interaction, swarms respond among each other.
- ✓ Principle of quality: It states that the quality factors of solutions are examined along with the simple solution.
- ✓ Principle of diversity: It states that the solution is obtained by searching entire search space not only the limited area.
- ✓ Principle of stability: It states that whenever there is an environmental change, the behavior of the swarm should be unchanged that is it should be stable.
- ✓ Principle of adaptability: It states that whenever there is an environmental change, the behavior of the swarms should also change.

II. LITERATURE REVIEW

Shi Cheng, Yuhui Shi, Quande Qin, and RuibinBai [2], presented a paper titled "*Swarm Intelligence in Big data Analytics*" which analyses the difficulties of big data analytics problems. It suggests the use of swarm intelligence in big data analytics. It is capable of solving large, multi-objective and dynamic problems.

Lim Kian Sheng et al. [3] proposed a paper on "*Multi-Objective Particle Swarm Optimization Algorithms – a Leader Selection Overview*". Here the paper clearly describes the Multi-Objective Optimization problem. Common features present in Multi-Objective Optimization problem and many performance measures are also explained. The paper also reviewed various Multi-Objective Particle Swarm Optimization Algorithms.

Ajith Abraham et al. [4] proposed a paper titled "*Swarm Intelligence Algorithm for Data Clustering*" which states the basic concepts of Swarm Intelligence and stresses the importance of particle swarm optimization and ant colony optimization algorithm. They also proposed a new fuzzy clustering algorithm, which depended on the variant variety of PSO.

Bing Xue and Will N. Browne [5] proposed a paper titled "*Particle Swarm Optimization for Feature Selection in Classification: a Multi-Objective Approach*" which performs searching of an optimized solution in an effective manner and obtains the sequence of non-dominated solution rather than a single search solution. The proposed algorithm assists the users to select their own preferred solution as per their requirement. The limitation of this paper is losing the swarm quickly, which limits its performance for feature selection.

Change Li and Shengxiang Yang [6] proposed a paper titled "*A Clustering Particle Swarm Optimizer for Dynamic Optimization*" which proposed a new algorithm to handle dynamic optimization problems. Though the proposed clustering algorithm is effective in creating sub swarms, it is still difficult to get the accurate sub swarms.

Dr. M.Seetha, G. Malini Devi, Dr.K.V.N.Sunitha [7] presented a paper titled "*An Efficient Hybrid Particle Swarm Optimization for Data Clustering"* which proposed an efficient hybrid method to solve the large sized fuzzy clustering problem. The proposed system had higher quality of solution as per the objective function value. They have told that the study can be extended further by combining other fuzzy clustering algorithms like those that K-means and then the best hybrid algorithm can be derived for large sized data set and using other measures like intensity and connectivity.

The paper by Q.Ni and Jading [8] proposed Dynamic Particle Swarm Optimization Algorithm, which is based on random topology. They also suggested a new PSO algorithm a logistic dynamic particle optimization that discusses the effectiveness of the random and design strategies of population topology.

III. METHODOLOGY

In computer science, particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. It solves a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best-known position, but is also guided toward the best-known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

PSO is originally attributed to Kennedy, Eberhart and Shi[1][2] and was first intended for simulating social behavior,[3] as a stylized representation of the movement of

organisms in a bird flock or fish school. The algorithm was simplified and it was observed to be performing optimization. The book by Kennedy and Eberhart[4] describes many philosophical aspects of PSO and swarm intelligence. An extensive survey of PSO applications is made by Poli.[5] [6] recently, a comprehensive review on theoretical and experimental works on PSO has been published by Bonyadi and Michalewicz.[7]

PSO is a metaheuristic as it makes few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. However, metaheuristics such as PSO do not guarantee an optimal solution is ever found. Also, PSO does not use the gradient of the problem being optimized, which means PSO does not require that the optimization problem be differentiable as is required by classic optimization methods such as gradient descent and quasi-Newton methods.

Initially upload a dataset sample and apply the concept of Jacquards distance on the dataset. Now on the obtained result apply the particle swarm optimization algorithm. Group of random particles are initialized first and by updating the generation optimal solution is obtained. In each iteration, evaluate the fitness function and update the pbest and gbest value. Pbest is the best solution so far and gbest is the value that is obtained by the particle swarm optimizer, obtained so far by any particle in the population. The Particle Swarm Optimization Algorithm which gives better seed selection by calculating forces on each particle due to another in each direction and the total force on an individual particle. The output of Particle SWARM OPTIMISATION ALGORITHM gives us our final clusters.

Step1: Start

Step2: Apply jaccards distance to the given

Dataset sample.

Step3: **for** each particle in the dataset **do**

Initialize particles with the random

Vector $x_i \sim U(b_{lo},b_{up})$

Initialize particles position to its initial

Position $p_i \quad x_i$

Step4: **if** $f(p_i) < f(g)$ **then**

Update the position of the particle to

$G \quad p_i$

Particles velocity is initialized to

$V_i \sim U(-|bup-b_{lo}| , |bup-b_{lo}|)$

Step5: Update the particles velocity by

$vi,d \quad \omega\ vi,d+ \varphi\ \Upsilon p(pi,d - xi,d) +$

$\varphi \quad \Upsilon_g(g_d - x_{i,d})$

$x_i \quad x_i + v_i$

Update the position of the particle

pi xi
g p_i

Step7: If termination criteria satisfies then

show the optimal solution.

Step8: Stop.
The output of this paper is the clusters of given dataset. Many big data analytics problems can be solved using swarm intelligence technique.

Here in this experiment particle swarm optimization algorithm is applied to the sample data set in order to obtain the clusters. Initially a fixed amount of document vector is selected by every particle swarm from the document collection as the centroid of the cluster. Now each particle is assigned with each document vector available in the dataset to the nearest centroid vector. Next as per the fitness function, fitness value will be calculated and finally new solution set is obtained by updating the velocity and particle position of each particle. We have to repeat the above steps until the maximum number of iteration is exceeded to the predefined number.

a) Before updating the position and velocity of the particle
b) After updating the position and velocity of the particle

IV. RESULT

We can calculate the accuracy using the following formula
Accuracy=total correctly classification/total sample classify

V. CONCLUSION

The main ambition of this paper is that it suggests the capability of swarm intelligence to with efficiency solve the

issues faced in big data analytics. With the appliance of swarm intelligence strong and effective algorithms are often developed to unravel big data analytics problem.

VI. REFERENCES

[1]. Sonu L al Gupta, Sofia Goel, Anurag Singh Baghel "An Approach to Handle Big Data Analytics Using Potential of SwarmIntelligence", International Conference on Computing for Sustainable Global Development, IEEE 2016.

[2]. Shi Cheng, Yuhui Shi, Quande Qin, and RuibinBai "Swarm Intelligence in Big data Analytics",14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013. Proceedings, pp 417-426.

[3]. Lim Kian Sheng et al. "Multi-Objective Particle Swarm Optimization Algorithms – A Leader Selection Overview", IEEE, DOI 10.5013/IJSSST.a.15.04.02.

[4]. Ajith Abraham et al. "Swarm Intelligence Algorithm for Data Clustering", In: Soft computing for Knowledge discovery and data mining, pp279-313, 2008.

[5]. Bing Xue and Will N. Browne "Particle Swarm Optimization for Feature Selection in Classification: A multi-objective approach," IEEE Transaction on Cybernetics, vol.43, no.6, pp.1656-1671, 2013.

[6]. Change Li and Shengxiang Yang "A Clustering Particle Swarm Optimizer for Dynamic Optimization", IEEE,978-1-4244-2959-2/09/$25.00-c2009.

[7]. Dr.M.Seetha, G. Malini Devi, Dr.K.V.N.Sunitha"An Efficient Hybrid Particle Swarm Optimization for Data Clustering"

[8]. International Journal of Data Mining & Knowledge Management Process (IJDKP) vol.2, No.6, November 2012.

[9]. Qing and J.Deng," Dynamic Particle Swarm Optimization Algorithm", based on Random Topology. The scientific world jounal.2013.

[10]. SonuLal Gupta, Sofia Goel, Anurag Singh Baghel "An Approach to Handle Big Data Analytics Using Potential of Swarm Intelligence", International Conference on Computing for Sustainable Global Development, IEEE 2016.

[11]. Shi Cheng, Yuhui Shi, Quande Qin, and RuibinBai "Swarm Intelligence in Big data Analytics",14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013. Proceedings, pp 417-426.

[12]. Lim Kian Sheng et al. "Multi-Objective Particle Swarm Optimization Algorithms – A Leader Selection Overview", IEEE, DOI 10.5013/IJSSST.a.15.04.02.

[13]. Ajith Abraham et al. "Swarm Intelligence Algorithm for Data Clustering", In: Soft computing for Knowledge discovery and data mining, pp279-313, 2008.

[14]. Bing Xue and Will N. Browne "Particle Swarm Optimization for Feature Selection in Classification: A multi-objective approach," IEEE Transaction on Cybernetics, vol.43, no.6, pp.1656-1671, 2013.

[15]. Changhe Li and Shengxiang Yang "A Clustering Particle Swarm Optimizer for Dynamic Optimization", IEEE,978-1-4244-2959-2/09/$25.00-c2009.

[16]. Dr.M.Seetha, G. Malini Devi, Dr.K.V.N.Sunitha"An Efficient Hybrid Particle Swarm Optimization for Data Clustering"

[17]. International Journal of Data Mining & Knowledge Management Process (IJDKP) vol.2, No.6, November 2012.

[18]. Q.Ni and J.Deng," Dynamic Particle Swarm Optimization Algorithm", based on Random Topology. The scientific world jounal.2013.