Advances in Dynamical Systems and Applications.
ISSN 0973-5321, Volume 16, Number 1, (2021) pp. 45-57

https://dx.doi.org/10.37622/ADSA/16.1.2021.45-57

WQPSO Method uses K-means-based Consensus Clustering in BigData

[1]Muthangi Kantha Reddy, [2]Dr.P. Srinivasa Rao,[3]Dr.E. Laxmi Lydia

[1]Director, Indo US Collaboration for Engineering Education (IUCEE), Hyderabad, India.

[2]Professor, Department of Computer Science and Systems Engineering, A.U College of Engineering, Andhra University, Andhra Pradesh, India.

[3]Professor, Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India.

Abstract

The consensus grouping expects to merge a few existing core segments into a coordinated set, which has generally been perceived for grouping heterogeneous and multi-source information. One can deduce from the strong and high-level performance of the usual grouping techniques draws by agreement in great consideration, and many efforts have been made to build this field. The K- means-based Consensus Clustering (KCC) changes the agreement grouping issue into a traditional K- means clustering with hypothetical backings and shows the favorable circumstances over the cutting edge techniques. Even though KCC acquires the benefits of K- means, it experiences assignment instantly. Also, the current system of aggregating arrangements isolates age and the combination of essential segments into two unrelated parties. To resolve the following two difficulties a Weighted Quantum Particle Swarm Optimization (WQPSO) with KCC is proposed. This paper proposes a WQPSO calculation with the weighted average of the best situation based on particle welfare estimates. Calculation WQPSO gives faster in the vicinity of mixing, the suites in a better harmony between the world and the neighborhood looking from the calculation so that it produces a great

performance. The proposed calculation of the WQPSO is well informed on some reference books and the contrasted and standard Particle Swarm Optimization (PSO). Similarly, in the grouping, there are many calculations of unassigned grouping that have been created such calculation is a KCC which is basic and direct. The Big Data Cluster contains the KCC calculation which is essentially used to decrease the length of the asset group.

Keywords: Consensus clustering, K-means-based Consensus Clustering, unsupervised clustering, Weighted Quantum Particle Swarm Optimization.

1. INTRODUCTION

PSO is a calculation based on the proposed stochastic population to be reused by the perceptive aggregation driving of certain creatures. In PSO, each molecule is considered a desired arrangement. All particles have a sense of well-being and speed, and they fly in a multidimensional chasing space by gaining chronicled data. It contains their memories of the best individual positions and information on the best situation worldwide in rallies during the investigative cycle. The PSO can be updated effectively and is from an economic IT perspective, and does not have many limits to change. Due to its predominance [1], PSO has rapidly evolved with applications taking care of certifiable improvement issues recently. In any case, PSO is easily taken into the optimum neighborhood, and premature assembly appears when it is applied to complex multi-modal issues. Many efforts have been made to improve the exposure of the PSO. Propelled by quantum mechanics and PSO's directional examination proposed a variation of PSO, which is referred to as Quantum action PSO (QPSO). Unlike the PSO, QPSO does not require velocity vectors for particles, and additionally has fewer limits to change, making it easier to run [2]. Since the QSO was proposed, it has attracted a great deal of attention and various variants of the QSO have been proposed to enhance the presentation of various perspectives. And efficiently applied to tackle a wide range of problems of constant improvement. As a general rule, most QPSO changes flow and reflux can be categorized into three classifications: improvement dependent on administrators of other development calculations, cross-search techniques, and useful strategies [3].

Even though these systems have improved the exposure of QPSO is regarding mounting speed and world optimal. It is relatively difficult to improve global prosecution capacity and speed up trade. In the QPSO example, both the best average individual position and the neighborhood attractor impact the exposure of the calculation. First, the previous one is essentially normal in terms of the best individual situation. All being equal, which overlooks the distinction of the impact of particles

with various welfare on directing the particles to look through world ideal arrangements. As a result, it does not help improve QPSO's global investigative capability. Also, the neighborhood attractor for a molecule can be obtained as the weighted amount of its own and best positions worldwide. It was discovered that there are not many improvements concentrating on neighborhood attractors in QPSO [4]. A swarm of quantum-action storytelling molecules with a Gaussian distributed local attractor point (GAQPSO) is proposed. In GAQPSO, the neighboring attractor depends on Gaussian diffusion whose average value is the first neighborhood attractor which is characterized by an exemplary QPSO. An Enhanced QPSO (EQPSO) depend on a novel treatment method for the nearby attractor is presented. In EQPSO, the neighborhood attractor is the weighted quantity of individual molecules and the best global positions. Use the capacity provided by the present and absolute quantities of the cycles as weight. This computational technique cannot filter the populace variation over time [5].

Subsequently, it is not useful for enhancing the global hunting capacity of the QPSO balance. Overall, variety is an important component of population-based promotion strategies because it impacts their exposure, and variety is strongly linked to the tradeoff of survey abuse. A wide variety encourages research, which is usually necessary for the course of the underlying emphasis of the advancement calculation. Low variety is characteristic of the abuse of a small area of the tracking space, desired during the last piece of the progress cycle. Observe the variety of QPSO populations to develop nearby attractors to manage the improvement of particulates. Subsequently improving the ability of computation to look through the global ideal and speeding up the rate of computation combination, this training is rarely advertised. In this paper, to adjust the world and neighborhood to the search abilities [6]. We propose a set of weighted coefficients that can recognize the well-being of the particles to determine the best individual mean position and a new method of treatment of the neighboring attractor. Also, a different kind of quantum-focused on enhancing the swarm of molecules with the weighted average best individual position and versatile neighborhood attractor is intended for mathematical rationalization. The exploratory results demonstrate that our proposed strategy is successful [7].

Even though the grouping of agreements has some advantages compared to the usual grouping strategies, it presents some difficulties. First of all, the grouping is a one-on-one task. Second, non-request ownership makes it difficult to fit clusters into various segments. Third, essential packages can have different group numbers. Third, the essential parcels may have distinctive group numbers. The tendency to the above difficulties in a structure assembled in their KCC calculation that modifies the grouping agreement to a (weighted) K-implies the grouping problem [8]. The changes give enormous benefits to the extent that the skill and the hypothetical help of

presenting KCC in all cases being capricious because K-implies is delicate to instantaneously. Moreover, the calculation does not produce the core package set. In this article, new clustering calculations thinking about the KCC issue, especially it depends on the voracious focus assignment in an extended segment include space. We are considering regulating the assignability of K-involves the introduction and age of fundamental segments in a collected system [9]. Stimulated by K-implies avaricious, a deeply competent variation of K-signifies that introduces K habitats with the past K 1 places K-implies introduction and age critical segment. In any case, eager K-implies, produces n segments with one certain group number, and just one segment has been chosen for the subsequent stage advancement, a kind of waste. Also, time's multifaceted nature becomes costly when n is huge.

2 RELATED WORKS

The purpose of bundling agreements is to find a solitary segment that fits into a few existing assignments, regardless of what one might expect. Normally, a public service capability is intended to quantify the arrangements between the fundamental segments and the last segment of the agreement at the segment level. On balance, the grouping of agreements can be formalized as a (combinatorial) improvement, problem with a given target capability, and it usually uses heuristics to discover hypothetical arrangements. Numerous calculations have been proposed to account for various target capabilities, including the Maximization of Expectations (MOE) calculation, nonnegative frame factorization, portion-based strategies, and reconstituted hardening. Among these techniques, a cutting-edge work drawn in much consideration, which uses a K-involves clustering to uncover the quadratic entropy dependent arrangement. Here, given a hypothetical system for KCC [10]. In recent years, KCC variations have been proposed to enhance this area, such as Disassemble-Assemble (DIAS), Spectral Ensemble Clustering (SEC), Entropy-based Consensus Clustering (ECC), and Infinite Ensemble Clustering (IEC). Even though these techniques have accomplished promising results, they experience all the negative effects of K-implies the introduction. A different set of strategies assesses proximity at the level of examples [11].

It characterizes a co-affiliation grid to count the occasions when two examples coexist in a similar group; this can be considered as another network of similarity. At this point, any graph segment technique can be applied to the co-affiliation grid to achieve the final result. A portion of these techniques incorporates co-affiliation grid-based strategies, relabeling and casting ballot techniques, locally versatile group-based techniques, hereditary calculation based techniques, otherworldly troupe bunching, and numerous different techniques. Remarkably, SEC joins these two types of

clustering strategies and shows that proximity at different levels may be convertible [12].

As K-implies, is sensitive to the introduction, a lot of efforts have been made to settle this test. The least difficult path is to execute many K-involves computations with instantaneously irregular and return one with the base target work. K-means++ applies a versatile examination technique to seed starting habitats K; this results in a calculation that carries out O (log K) - serious calculation with the ideal grouping. Also, K-implies test a few focuses each time with various runs and leads a weighted K-implies on these examined focuses to deliver K bunches for reinstatement. Avaricious K-implies, is a gradual way to deal with powerfully adding every medium in turn by a deterministic pursuit around the world. Different strategies incorporate the technique based on the closest neighbor, the strategy based on the proliferation of partiality, and K-involves recursive. Although some reviews have been conducted on the K-implies aggregation agreement instantaneously, it is the main study to fully address the introduction of KCC assignment [13].

Motivated by the voracious area system of K-involves voracious, we strengthen the excesses of insatiable K-means and KCC to plan our new calculation, GKCC. The fundamental thought uses the transient segments found by means K varies to characterize the layout of the essential plots and for the last step of combining. Furthermore, the greedy technique of voracious K-involves solving the question of introduction that infests K-involves methodologies. Lastly, a hypothetical start-up aid is given as a blunder limit [14].

3 METHODOLOGY

3.1 Big Data Analytics

The advanced web and globalization have opened by open organizations same registration information, Big Data turns into dreaded activity. The ability to get (catch), handle and separate (look), also gives a huge advantage by extending the organization's ability to meet dynamic financial circumstances and customer needs. On the way out, the customer worker's plan for Big Data is shown in Figure 1. The association of knowledge and business exam (BI&A) in three classes depending on business information was planned by Chen, Chiang, and history in the year 2012. The first form of a survey that uses the DBMS-based substance for work (BI&A 1.0) and also uses different data storage devices, ETL, OLAP. The next BI&A 2.0 rendering is an electronic programming content in which the association of information is optional, primarily on information retrieval, web assessment online media review, and transient space review. The third and most refreshed rendering capabilities of a

versatile, sensor-based substance is BI&A 3.0. It emphasizes portable representation, associations, and information review. The second and third forms must mix on an immense measure of information, speed, and also the different types of volume information which are grouped.

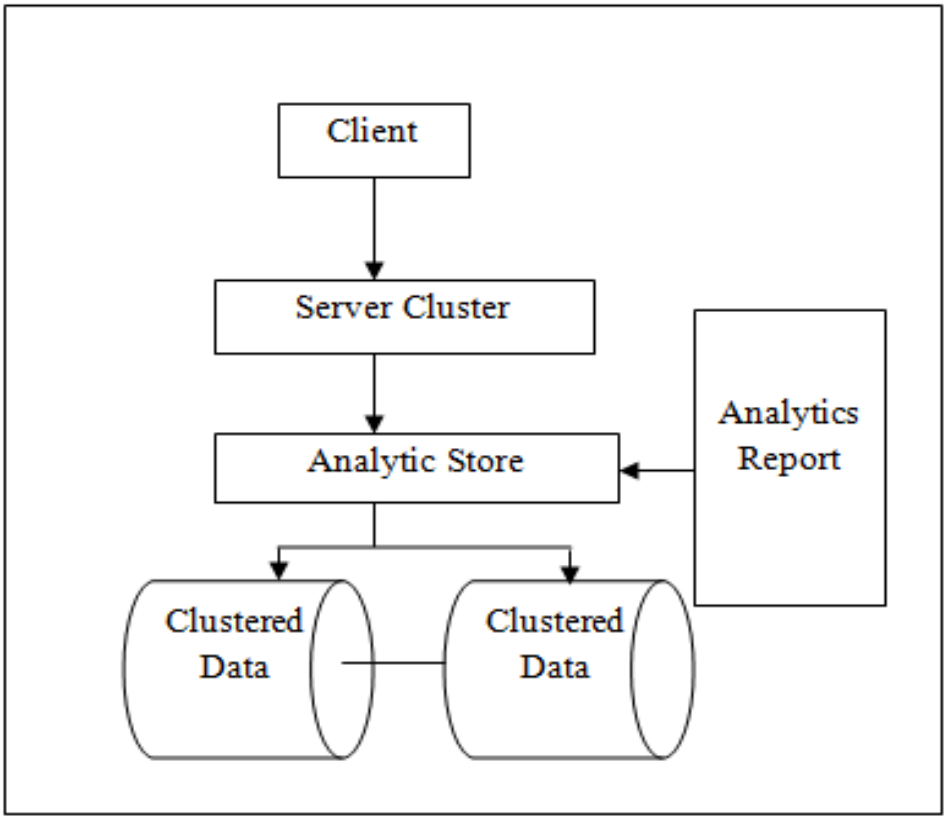

Figure 1: A Consensus Cluster Architecture for Big Data

The analytical design of Big Data portrays according to tradition the extremely even and scattered volume other than the planning framework given by Hadoop HDFS and MapReduce. Rather than a certain conviction that data storage and massive information centers around. Again, the information store may be a hotspot of data for compound Hadoop races, at the same time using the extremely equal limits of two structures. Incremental regional data from GPS or phones are attached to noteworthy data, information storage for continued understanding. Sponsors should raise individual customer-centric questions based on zone data and customer profiles.

3.2 Particle Swarm Optimization

It is a calculation based on the proposed stochastic population, which is inspired by

the overall canny conduct of certain creatures, e.g., flocks of poultry or schools of fish. Future responses for PSO are referred to as particulate matter. The developments of particles are driven by their own most popular position called pbest and the most popular position of the whole multitude called gbest. An idea for enhancing nonlinear capabilities using the molecule swarm strategy is presented. The progress of some norms is sketched out, and the use of one of the ideal models is mentioned. Reference tests of world vision are depicted, and applications, including the advancement of non-linear capability and the preparation of neuronal organization, are proposed. The links between the rationalization of the molecular swarm and both false life and hereditary stones are depicted. PSO carries out the research using a multitude of particles which refreshes the cycle in emphasis. To find the ideal arrangement, each molecule travels towards the path to its best position (pbest) and the best position (gbest) in the world in the multitude.

$$X_i(t+1) = X_i(t) + V_i(t+1) \text{---------- (1)}$$

Where X_i (t+1) represents the particle area I at the $(t+1)^{th}$ cycle, V_i (t+1) indicates the gbest and pbest i at the $(t+1)^{th}$ emphasis.

3.3 Weighted quantum-behaved particle swarm optimization

As mentioned above, in the PSO, the best mean position m is accustomed to evaluating the estimation of L, which makes the calculation more efficient than the one proposed. We can see that the best average position is normal as far as the best individual situation is concerned. All being equal, which implies that each molecule is considered to be approachable and applies a similar effect on my estimation. The way of thinking about this strategy is that the Mainstream Though, that is to say the best position m, decides on the further expansion or innovation of the molecule. The significance of the dominant thought as the medium of the best individual positions is quite reasonable. The weighted average position, in the same manner, is nevertheless somewhat strange, contrasting, and developing a social culture in the authentic world. For a certain thing, even though the whole social life form determines the dominant thought, it is not appropriate to consider each equivalent part. The elitists are playing a more important role in cultural advancement. From this perspective, when we plan another control strategy for the QPSO in this document, m is swapped for a weighted average position. The main question is to decide whether a molecule is elitist or not, or to state it with precision, how to evaluate its significance in the calculation of m

estimation. It is normal, as in other development calculations, for us to associate elitism with the well-being estimates of particles as shown in Figure 2.

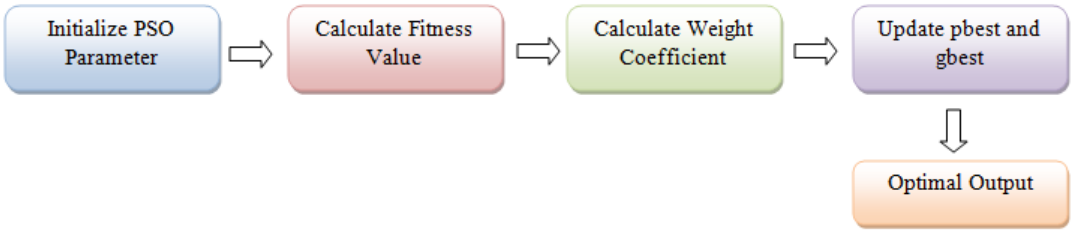

Figure 2. WQPSO Implementation steps

The more aptitude, the more important the molecule. By officially portraying it, we can classify the molecule in the descending demand as indicated by their well-being estimate first. At this point, each molecule loses a weight coefficient that decreases with the position of the particle, meaning that the closer the best arrangement is, the higher its weight coefficient. The best average position m, that way, is determined as.

$$p(s) = p1(s),p2(s),....,pn(s)= 1/p\left(\sum_{i=1}^{p}(f_{ai} - f_{di})^2(t)\right)\text{----------(2)}$$

Where a is the weight coefficient and di is the dimension coefficient of every particle, s is the population size. In this paper, the weight coefficient for each particle decreases linearly from 1.5 to 0.5.

Algorithm Steps for WQPSO

1. Initiate data size settings.
2. Report the particle population as per positions.
3. Determine the best optimization values (gbest).
4. Calculate the population of particulates using the weight factor.
5. Keep the particle size coefficient m up to date.
6. Update status position value fi.
7. If the specified condition is not met, continue with step 3.

4 PERFORMANCE EVALUATION

4.1 Time Computation

For n^{th} Cluster resource, the time computation factor is shown as:

$$T(n) = e_n + i_n + r_n \text{----------(3)}$$

The Execution Time of the nth cluster in i^{th} Cluster resource also Retrieving Time of nth cluster in n^{th} Cluster resource in Figure 3.

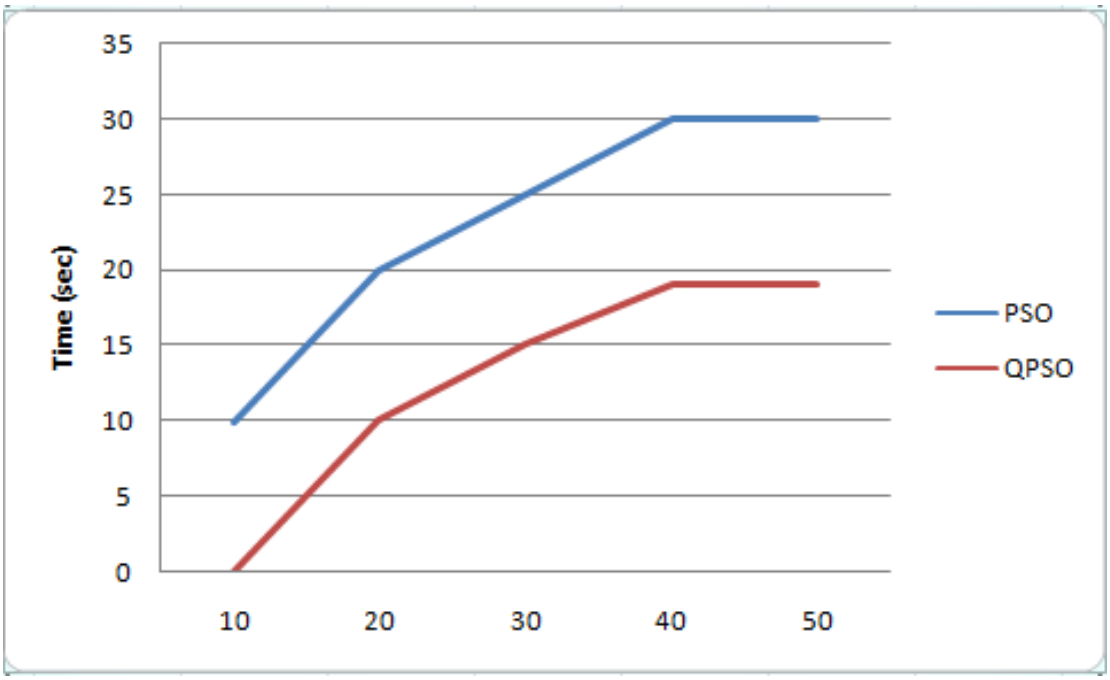

Figure 3. Time Consumption vs Number of Clusters

4.2 Load Computation

The n^{th} Cluster resource at nth cluster load computation results is calculated and shown in Figure 4.

$$\text{Load computation} = \frac{\text{no.of relevant data extracted}}{\text{total no.of data extracted}} * 100\% \text{----------(4)}$$

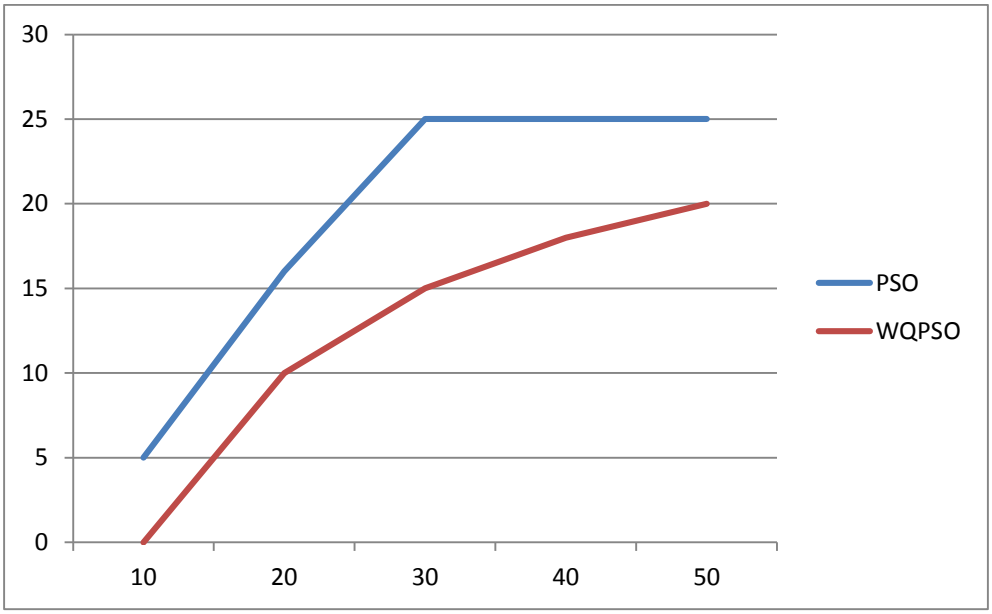

Figure 4. Load Computation vs Number of Clusters

4.3 Energy Computation

The total energy requirement for not cluster in i^{th} cluster resource is shown below in Figure 5.

$$E(n) = i_n + e_{(n,x)} \text{ ----------(5)}$$

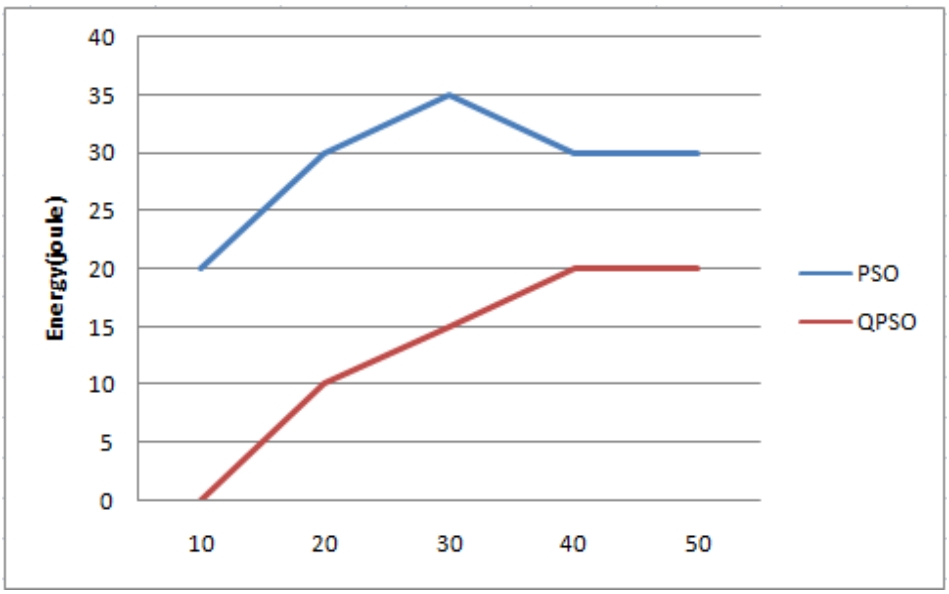

Figure 5. Energy Consumption in Joule

4.4 System Reliability Vs. Number of Clusters

The general framework dependability is summed up as the normal of each bunch present in the organization, i.e unwavering quality of each group is considered shown in Figure 6.

Measurably characterized as follows:

$$Reliability\ rate = \frac{no.of\ \ data\ clutered}{total\ no.of\ data\ in\ database} * 100\% \quad \text{--------(6)}$$

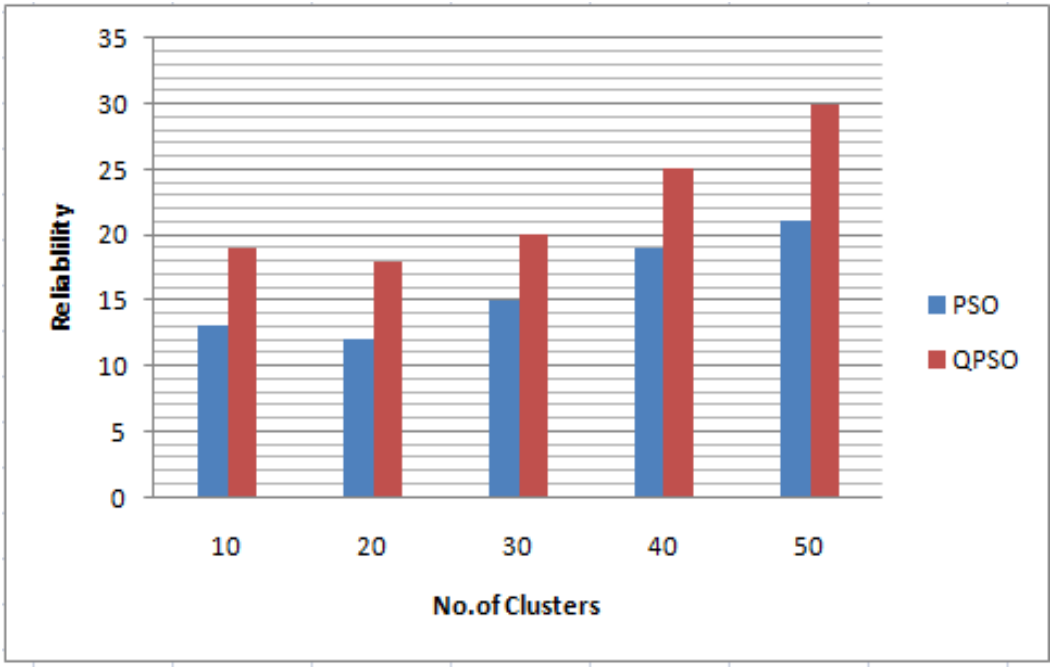

Figure 6. Reliability rating vs the number of clusters

5 CONCLUSIONS

A variant of the PSO to be specific WQPSO is offered to all while improving the continued performance of WQPSO and gaining an ideal large capacity around the world. Initially, a weight limit is known to recognize the distinction between the impact of particles and various well-being. It is used to obtain a weighted average of the best individual position of the population. Also, a direct mixture of the most popular position of the molecule and the entire set is intended to shape the versatile neighboring attractor. Using the number of square deviations from PM welfare estimates such as the right mixture coefficient. The goal of this progress is to improve the global survey in the early stages of rationalization. To induce particles to fuse toward the global optima toward the end of the hunt. Finally, the proposed calculation of the WQPSO was tried out over twelve reference works and contrasted with the

fundamental artificial bee colony and the other four variations of the PSO. The results of the tests show that WQPSO works better than the strategies analyzed in the set of reference capabilities regarding global survey capacity and faster assembly rate. We infer that distinguishing the well-being of particles to depict the best weighted average individual position. Observing the variety of the PSO population to build a multipurpose neighborhood attractor to control particulate enhancement, is feasible. Even though the proposed WQPSO showed an unparalleled performance in the results of the trial. Detailed in the preceding sub-sections, it is just suitable for unconstrained questions in the constant survey space. Other changes should never really broaden the relevance of the proposed WQPSO to a broader category of improvement issues, including discrete, multi-purpose, mandate, and powerful advancement issues. Similarly, the Big Data Cluster contains the consensus clustering calculation based on K-averages, which is mainly used to reduce the length of the asset group.

REFERENCES

[1] Chen, S. (2019). Quantum-Behaved Particle Swarm Optimization with Weighted Mean Personal Best Position and Adaptive Local Attractor. Information, 10(1), 22.

[2] Khan, M., Huang, Z., Li, M., Taylor, G. A., & Khan, M. (2017). Optimizing hadoop parameter settings with gene expression programming guided PSO. Concurrency and Computation: Practice and Experience, 29(3), e3786.

[3] Sadasivam, G. S., & Selvaraj, D. (2010, December). A novel parallel hybrid PSO-GA using MapReduce to schedule jobs in Hadoop data grids. In 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC) (pp. 377-382). IEEE.

[4] Latchoumi, T. P., Ezhilarasi, T. P., & Balamurugan, K. (2019). Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 1(10), 1137.

[5] Lydia, E. L., & Swarup, M. B. (2016). A Disparateness-Aware Scheduling using K-Centroids Clustering and PSO Techniques in Hadoop Cluster. International Journal, (02).

[6] Wai, E. N. C., Tsai, P. W., & Pan, J. S. (2016, November). Hierarchical PSO clustering on mapreduce for scalable privacy preservation in big data. In International Conference on Genetic and Evolutionary Computing (pp. 36-44). Springer, Cham.

[7] Latchoumi, T. P., & Sunitha, R. (2010, September). Multi agent systems in distributed datawarehousing. In 2010 International Conference on Computer and Communication Technology (ICCCT) (pp. 442-447). IEEE.

[8] Kamel, N., Ouchen, I., & Baali, K. (2014). A sampling-pso-k-means algorithm for document clustering. In Genetic and evolutionary computing (pp. 45-54). Springer, Cham.

[9] Ranjeeth, S., Latchoumi, T. P., & Victer Paul, P. (2019). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications. https://doi. org/10.2174/2666255813666191116150319.

[10] Lam, Y. K., Tsang, P. W. M., & Leung, C. S. (2013). PSO-based K-Means clustering with enhanced cluster matching for gene expression data. Neural Computing and Applications, 22(7-8), 1349-1355.

[11] Wu, J., Liu, H., Xiong, H., Cao, J., & Chen, J. (2014). K-means-based consensus clustering: A unified view. IEEE transactions on knowledge and data engineering, 27(1), 155-169.

[12] Loganathan, J., Janakiraman, S., & Latchoumi, T. P. A Novel Architecture for Next Generation Cellular Network Using Opportunistic Spectrum Access Scheme. Journal of Advanced Research in Dynamical and Control Systems,(12), 1388-1400.

[13] Li, X., & Liu, H. (2018). Greedy optimization for K-means-based consensus clustering. Tsinghua Science and Technology, 23(2), 184-194.

[14] Loganathan, J., Latchoumi, T. P., Janakiraman, S., & parthiban, L. (2016, August). A novel multi-criteria channel decision in co-operative cognitive radio network using E-TOPSIS. In Proceedings of the International Conference on Informatics and Analytics (pp. 1-6).