International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248
Volume: 4 Issue: 1 303 – 307

Analysis of Cloud Storage Issues in Distributed Cloud Data Centres by Parameter Improved Particle Swarm Optimization (PIPSO) Algorithm

Baalamurugan K.M
Research Scholar
Department of Computer Science & Engineering
Annamalai University
Chidambaram, India
baalaresearch@outlook.com

Dr. S. Vijay Bhanu
Assistant Professor
Department of Computer Science & Engineering
Annamalai University
Chidambaram, India
vbhanu22@yahoo.in

Abstract — Cloud Computing environment provides several on demand services for users. The recent problem and important issues in cloud computing is to optimize cloud storage over distributed data centers. The cloud storage optimization problem is considered as one of most challenging task. To overcome storage issues, the Parameter Improved Particle Swarm Optimization (PIPSO) technique is proposed. This paper presents the objective and the requirements of the distributed cloud storage issue based on the current topology. PIPSO technique manages each other between two server centers. The simulation has been done in MATLAB programming software. The results show that proposed technique achieves better results when compared with the other existing methods.

Keywords **-** *Cloud Computing, Optimization, PSO, GA, PIPSO, Distributed Data Centres.*

*****

I. Introduction

The need for storage resources and computing resources are growing faster in recent years. Therefore, cloud computing enters into play due to high performance computing services and facilities provide to user [1]. Cloud computing is one of the most interesting research topics in recent years. The primary objective of the cloud computing is to make use of resources like processing, storage and system resources [2]. The important features of cloud data centres are backup data and cloud services, which are used by most of the companies in recent decades [3].

In distributed cloud computing, modern server centre is one of the key elements. The advanced local server centre cannot afford major capacities including processing, storage and system capacity. The security issues in [4] proposed cloud computing which comprises validation of high versatility, adaptability and single interface.

The most essential research topics in cloud computing is storage of information between distributed servers centres. Furthermore, the data's stored contains various types of information's. The test datasets are obtained from UCI machine learning repository [5]. The disturbed cloud storage issues can be generally named as discrete optimization problem. It is obvious that the client access latency can be reduced with the number of replicas increased and every client demands to access its data set from a replica that is as close as possible in order to minimize its delay. However, there are at least two challenges while replicating all data sets to all data centres can ensure low-latency access [6]. First, the system cannot offer unlimited storage resources, and that approach is costly and may be inefficient for the extra storage resource consumption. Second, the problem is more complicated when the data set may be updated, and the more the replicas are in the system, the higher the update cost will be. Thus, data set replicas need to be carefully placed to avoid unnecessary expense. And the replicas number and their store placements may have a profound impact on the optimal replicas distribution in a balance way from cost-effective view. The resulting trade off between the number of replicas and the data set delay maps precisely to the replicas placement problem.

The paper further sorted as follows, in section II previous research works on distributed cloud storages issues are presented. The fundamental structures of distributed server centres are depicted with problem formulation in Section III. Section IV describes about the constraints of proposed methodology. The PISPO technique is formulated in Section V. Simulation results are explicated in section VI. Section VII concludes the outcomes of the data centres.

II. Related Work

This section discuss about the previous research approaches related to cloud storage problems. Le et al. [7] has proposed the studies related to the advantaged of using green energy during data centres operations. The study of multi data centre framework services were introduced in [8]. In [9] provides a solution to place data to each data centre such that to minimize the data movement. This approach gives better solution to the application streaming the videos. Based on the parameter, data are placed to the near location of the user. Microsoft team [10] gives solution for data placement to reduce the data movement between geo-distributed data

ISSN: 2454-4248
303 – 307

centre. Based on server logs the decision to place data in server is granted. [11] and [12] proposes genetic algorithm based placement strategy for data centre resizing.

Optimization formulation has been introduced to reduce the cost and power consumptions in cloud is achieved [13]. In [14] has introduced model for reduced carbon emission. Reducing the cost for the usage of virtual machine usage has not received much attention. Only few measures to optimize multi objective in cloud survive. [15] has introduced minimized cost, latency and bandwidth approaches for cloud.

Clustering the data using bio-inspired algorithm has been proposed in [16] for easy computation of data's stored in cloud data centres. In most of the cloud environment the cloud data centres are distributed geographically. In [17] and [18] has extended the research in distributed data centre architecture.

In [19] gives a brief overview of data migration to cloud. The work also explains different steps to be carried out while transferring large data to cloud. The difficulties faced in data migration is analysed in [20]. To transfer large amount of data along with cache method is discussed in [21].

III. PROBLEM FORMULATION ON DISTIBUTED CLOUD DATA CENTRES

This section explains about the problems of storing data across distributed data centre locations. The data stored across locations contains structured and unstructured information's. The data has the properties [22] as follows

$$D_i = \{p_i, q_i, x_i, y_i\} \quad (1)$$

Where,
P_i - ID of Data
Q_i – the name depicting i^{th} Data
X_i and Y_i – Longitude and Latitude of data

The storage of data form one centre location to another centre location by limiting the transmission length is very important. The distributed data centre set S is defined by,

$$S = \{s_1, s_2, s_3, \ldots\ldots, s_n\} \quad (2)$$

Where n is the number of distributed data centres. Figure 1 shows the model of distributed data centres at various locations.

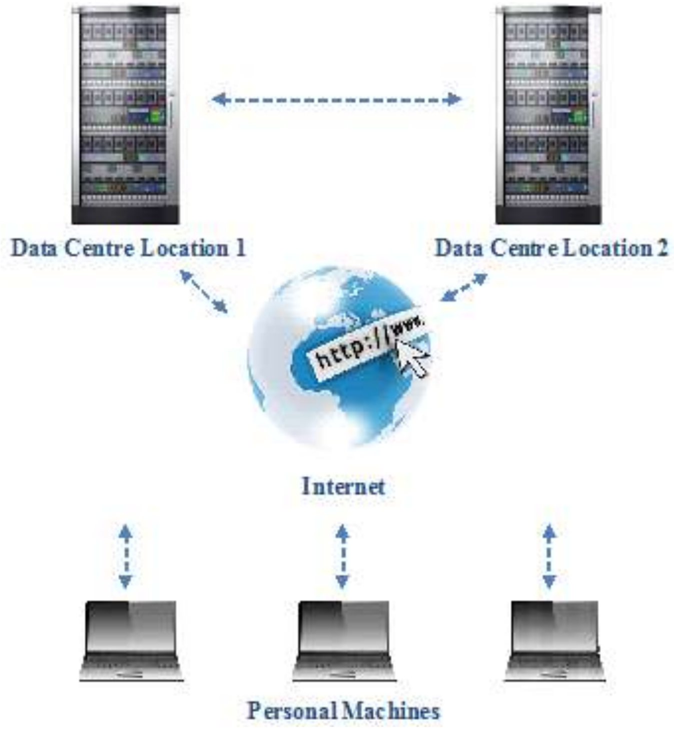

Figure 1. Distributed Data Centre

The Data centre contains Virtual Machine, software, storages and program components. The local machine can access any data centre location to create virtual machine and store data on the storage pool. The whole data sets of the distributed data centres is calculated by,

$$Z = \sum_{i=1}^{n} z_n \quad (3)$$

IV. CONSTRAINTS OF THE PROPOSED METHODOLOGY

To demonstrate the cloud storage problem with proposed work, various assumptions are referred form [22]. The data centres have the same prototype (i.e.) similar to each other by components. The transmission and delay loss are not considered in the process. The two adjacent data centres are calculated by longitude and latitude for transmission of data.

The node set S physically denotes the data centre as described in equation (2). The distributed data centres are denoted by *S'* given by,

$$S' = \{s'_1, s'_2, s'_3, \ldots\ldots, s'_n\} \quad (4)$$

The edge set E refers to the direct topology between two nodes

$$e = \begin{cases} 1 \\ 0 \end{cases} \quad (5)$$

The distributed cloud storage problem can be typically classified by the discrete optimization problem. The distance between the node set can be calculated from Euclidean distance between the data centre is described by,

$$D(s_i, s'_j) = \sqrt{(x_{si} - x'_{sj})^2 + (y_{si} - y'_{sj})^2} \quad (6)$$

V. PROPOSED METHODOLOGY

In order to deal with ideal distributed cloud storage issues of the distributed cloud data centers the PISPO technique is used which derived from the continuous PSO algorithm [23], [24], [25], [26], [27]. The PIPSO strategies has been used in various research fields [28] and [29]. To utilize the current PIPSO techniques, the mapping function from the consistent space to discrete space can be basically accomplished. This can be done by arranging the operator of constant esteems. The PIPSO technique consists of mapping order between the node S and S'. The parameters in every particle can figure the target function and should satisfy the requirements. The basic steps of PIPSO technique can be generally finished up by the following steps.

PSO algorithm explores for global optimum solution in a given N-dimensional problem by collaborating with particles in a swarm. Each particle or individual in a swarm has the characteristics of position and velocity. Mathematically, position 'x' and velocity 'v' of a particle can be represented as feasible solution of the problem and its step length for next iteration respectively. For a given N-dimensional mathematical problem and 'm' number of particles, the position and velocity of the i^{th} particle is denoted as $x_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]$ and $v_i = [v_{i1}, v_{i2}, \ldots, v_{iN}]$ respectively. At the end of each iteration, the best position of i^{th} particle compared with previous iterations is registered as local best solution and denoted by $p_i = [p_{i1}, p_{i2}, \ldots, p_{iN}]$. The overall best position among all particles in a population is considered as global best position $p_g = [p_{g1}, p_{g2}, \ldots, p_{gN}]$. The velocity and position update of i^{th} particle for next iteration can be computed by using particle's current velocity and its distance between local best position and global best position, the mathematical expression for i^{th} particle's velocity and position updates are as follows,

$$v_{id}^{t+1} = \omega . v_{id}^{t} + \varphi_1 (p_{gd} - x_{id}^{t}) + \varphi_2 (p_{id} - x_{id}^{t}) \quad (7)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (8)$$

where, t is the iteration index; ω is the inertia weight; d is the dimension index; $\varphi_1 = c_1 r_1$ and $\varphi_2 = c_2 r_2$; c_1 and c_2 are two positive acceleration coefficients called social and cognitive agents respectively; both r_1 and r_2 are random numbers uniformly distributed in between 0 and 1.

By extending the conception of PSO, Parameter Improved Particle Swarm Optimization (PIPSO) algorithm has been developed by updating PSO parameters such as inertia weight, social and cognitive agents at each iteration. This improved version of PSO algorithm can attain both local and global solutions and can take minimum number of iterations to convergence for global optimal solution of a problem. The parameters ω, c_1 and c_2 can be calculated from the following equations,

$$\omega_1 = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{t_{max}} \times t \quad (9)$$

$$\omega = \omega_{min} + \omega_1 \times r_3 \quad (10)$$

$$c_1 = c_{1max} - \frac{c_{1max} - c_{1min}}{t_{max}} \times t, \quad (11)$$

$$c_2 = c_{2max} - \frac{c_{2max} - c_{2min}}{t_{max}} \times t, \quad (12)$$

where ω_{min} and ω_{max} are minimum and maximum weights; c_{1min} and c_{1max} are minimum and maximum cognitive factors; c_{2min} and c_{2max} are minimum and final social factors; t_{max} is maximum iteration; r_3 are random numbers uniformly distributed in between 0 and 1.

VI. SIMULATION RESULTS

Keeping in mind the end goal to exhibit the effectiveness and the execution of taking care of with the distributed cloud storage issue by the PIPSO technique, numerical outcomes mostly focus on the distributed cloud storage issue of test dataset. Because of the randomness of the PIPSO technique, the suboptimal solution of the distributed cloud storage issue conceivably isn't equivalent to each other, particularly when the quantity of server centres is for the most part bigger than 15. Consequently, the target of this segment is to talk about the ideal or problematic methodology of the distributed cloud storage issue under various goals and diverse quantities of server centres, reducing the small transmitting length. On account of the limitations in this issue, the new particle ought to be fulfilled three limitations at each transformative procedure. The m-script has been created and executed in MATLAB simulation software on Intel Pentium quad core processor with 8GB RAM. We have considered 2 data center location, which has data capacity of same size stored on each. Each data center is connected to each other. The convergence graph and correlation of fitness function for GA, PSO and proposed PIPSO techniques are depicted in Figure 2 and 3 separately. Table 1 demonstrates the correlation result of data centre location obtain from GA, PSO and proposed PIPSO.

PIPSO parameters used for analysis are,

- Number of particles, m = 10
- Maximum number of iterations, $t_{max} = 200$
- Acceleration coefficients, $c_{1min} = c_{2min} = 1.5$ and $c_{1max} = c_{2max} = 2.2$,
- Inertia weight, $\omega_{min} = 0.3$ and $\omega_{max} = 1.2$

VII. CONCLUSION

The distributed cloud storage issue of server centres are examined and analyzed by the PIPSO technique. The outcomes acquired from the proposed PIPSO technique proves that the proposed procedure has better execution when compared against other existing techniques.

TABLE I. COMPARISON RESULT OF DATA CENTRE LOCATION OBTAIN FROM GA, PSO AND PROPOSED PIPSO

GA		PSO		PIPSO	
Location 1	Location 2	Location 1	Location 2	Location 1	Location 2
0.9462	0	0.9253	0	0.9756	0
0.6898	-0.1835	0.6548	0.0761	0.756	-0.0014
0.7116	0.039	0.7208	0.1373	0.8098	0.0594
0.6137	0.0227	0.6361	0.0076	0.739	0.0244
0.6382	0.1283	0.5696	0.1557	0.7345	0.0607
0.6062	0.0105	0.5904	0.1983	0.6429	0.0699
0.4954	0.0327	0.7067	0.2106	0.6241	-0.1091
0.5605	-0.0339	0.4924	0.0656	0.63	0.1507
0.5016	0.0666	0.4937	0.0267	0.6778	-0.0631
0.6128	-0.125	0.4507	0.0559	0.7064	-0.0843
0.5401	0.0456	0.5276	-0.0354	0.5311	-0.0097
0.4436	-0.1807	0.5461	-0.1882	0.5953	-0.0442
0.5883	-0.1591	0.4539	-0.0212	0.5197	-0.0111
0.6172	-0.099	0.6757	-0.1716	0.6525	0.0459
0.4983	0.0212	0.4738	0.0097	0.5256	-0.0532
0.4469	-0.0042	0.3736	0.0607	0.6445	0.0773
0.4629	-0.0058	0.4663	-0.0028	0.438	0.0758
0.9935	0	1	0	1	0
0.9118	-0.3089	0.7612	0.1318	0.6326	0.1087
0.6112	0.2104	0.5345	-0.1104	0.6838	0.157
0.5155	0.4722	0.682	-0.0512	0.4933	-0.0475
0.6464	0.0147	0.7126	0.3481	0.571	0.3434
0.6607	0.2754	0.2236	0.419	0.2745	0.3133
0.5325	0.119	0.5882	-0.0612	0.3227	0.316
0.313	0.0185	0.2077	0.0035	0.1896	0.3906
0.4723	0.0565	0.4696	0.0592	0.407	0.1649
0.3163	0.023	0.2942	0.1189	0.4189	-0.014
0.3062	0.1525	0.3589	-0.0141	0.0338	0.3546
0.6399	-0.3702	0.5882	0.1191	0.5137	-0.4287
0.3633	-0.1395	0.2914	0.1605	0.4131	0.1656
0.5611	-0.2228	0.6855	-0.3803	0.5127	0.3213
0.5206	0.0807	0.0058	-0.1769	-0.0329	-0.2174
0.4811	-0.0238	0.1393	-0.0109	0.5454	0.1204
-0.1045	-0.1574	0.3205	0.375	0.3458	0.1829

The single server centre and the current correspondence system of the distributed server centres are acquainted agreeing with the current existing data sets. At that point, the distributed cloud storage issue, including two goals and three limitations, is basically one of the discrete issues with a few requirements. Keeping in mind the end goal to handle with the previously mentioned issue, the PIPSO technique is used to coordinate the set node S and the set node S', finally achieving the small transmitting length. The future work is to discuss that the corresponding results can be employed to the development of the distributed server centres. Also, another fascinating distributed cloud storage issue is that one centre data is progressively doled out to the contiguousness server centre to accomplish the small transmitting length, when the storage information volume of each server centre focus isn't equivalent to each other.

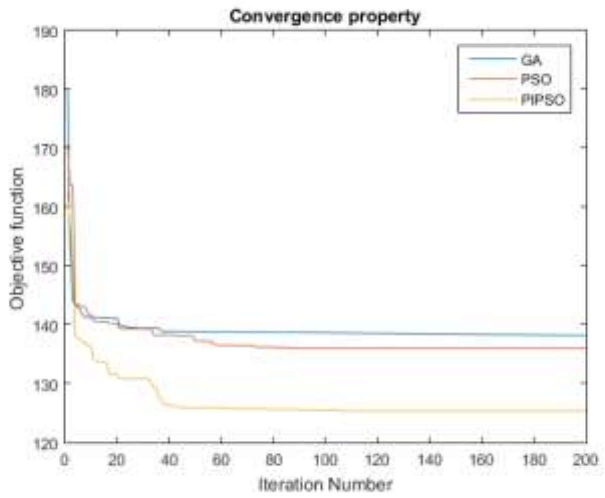

Figure 2. Convergence Characteristic of GA, PSO and PIPSO

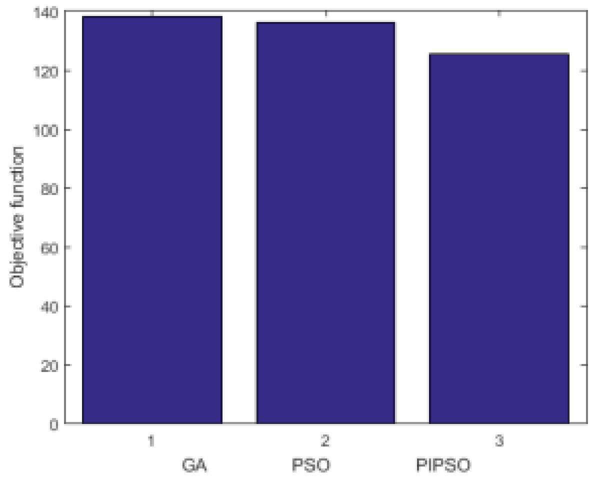

Figure 3. Comparison of fitness function for GA, PSO and PIPSO

ISSN: 2454-4248
303 – 307

REFERENCES

[1] A. M. Manasrah and H. Ba Ali, "Workflow Scheduling Using Hybrid GA-PSO Algorithm in Cloud Computing," *Wirel. Commun. Mob. Comput.*, vol. 2018, p. 16, 2018.

[2] M. N. O. Sadiku, S. M. Musa, and O. D. Momoh, "Cloud Computing: Opportunities and Challenges," *IEEE Potentials*, vol. 33, no. 1, pp. 34–36, Jan. 2014.

[3] Z. Jin, G. Xu, Y. Li, and P. Liu, "A novel cloud scheduling algorithm optimization for energy consumption of data centres based on user QoS priori knowledge under the background of WSN and mobile communication," *Clust. Comput.*, vol. 20, no. 2, pp. 1587–1597, Jun. 2017.

[4] G. Yan, D. Wen, S. Olariu, and M. C. Weigle, "Security challenges in vehicular cloud computing," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 284–294, Mar. 2013.

[5] "UCI Machine Learning Repository." [Online]. Available: https://archive.ics.uci.edu/ml/index.php. [Accessed: 04-Feb-2018].

[6] X. Wu, "Data Sets Replicas Placements Strategy from Cost-Effective View in the Cloud," *Sci. Program.*, vol. 2016, pp. 1–13, 2016.

[7] K. Le, R. Bianchini, J. Zhang, Y. Jaluria, J. Meng, and T. D. Nguyen, "Reducing electricity cost through virtual machine placement in high performance computing clouds," in *2011 International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, 2011, pp. 1–12.

[8] K. Le, O. Bilgir, R. Bianchini, M. Martonosi, and T. D. Nguyen, *Managing the Cost, Energy Consumption, and Carbon Footprint of Internet Services ∗* . .

[9] Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, and F. C. M. Lau, "Scaling Social Media Applications Into Geo-Distributed Clouds," *IEEEACM Trans. Netw.*, vol. 23, no. 3, pp. 689–702, Jun. 2015.

[10] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, and H. Bhogan, "Volley: Automated Data Placement for Geo-Distributed Cloud Services," in *NSDI*, 2010.

[11] Z. Er-Dun, Q. Yong-Qiang, X. Xing-Xing, and C. Yi, "A Data Placement Strategy Based on Genetic Algorithm for Scientific Workflows," 2012, pp. 146–149.

[12] L. Gu, D. Zeng, P. Li, and S. Guo, "Cost Minimization for Big Data Processing in Geo-Distributed Data Centers," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 3, pp. 314–323, Sep. 2014.

[13] J. Cao, K. Li, and I. Stojmenovic, "Optimal Power Allocation and Load Distribution for Multiple Heterogeneous Multicore Server Processors across Clouds and Data Centers," *IEEE Trans. Comput.*, vol. 63, no. 1, pp. 45–58, Jan. 2014.

[14] J. Doyle, R. Shorten, and D. O'Mahony, "Stratus: Load Balancing the Cloud for Carbon Emissions Control," *IEEE Trans. Cloud Comput.*, vol. 1, no. 1, pp. 1–1, Jan. 2013.

[15] A. Ruiz-Alvarez and M. Humphrey, "A Model and Decision Procedure for Data Storage in Cloud Computing," in *2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)*, 2012, pp. 572–579.

[16] K. M. Baalamurugan and S. V. Bhanu, "An efficient clustering scheme for cloud computing problems using metaheuristic algorithms," *Clust. Comput.*, Jan. 2018.

[17] "MEC-IDC: joint load balancing and power control for distributed Internet Data Centers - Semantic Scholar." [Online]. Available: /paper/MEC-IDC-joint-load-balancing-and-power-control-for-Rao-Liu/27ad61cc0dcb082a4fe5b411c7d108ecb71403cf. [Accessed: 04-Feb-2018].

[18] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi-Electricity-Market Environment," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.

[19] D. Jiaman, W. Sichen, D. Yi, and J. Lianyin, "A Dynamic Migration Method for Big Data in Cloud," in *2016 17th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2016, pp. 95–98.

[20] P. Nahar, A. Joshi, and A. Saupp, "Data Migration Using Active Cloud Engine," in *2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, 2012, pp. 1–4.

[21] K. W. Yan, N. M. Perumal, and T. Dillon, "Data migration ecosystem for big data invited paper," in *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 2013, pp. 189–194.

[22] J. Liu, T. Shi, and P. Li, "Optimal cloud storage problem in the distributed cloud data centers by the discrete PSO algorithm," in *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 156–163.

[23] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Inf. Process. Lett.*, vol. 85, no. 6, pp. 317–325, Mar. 2003.

[24] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[25] F. van den Bergh and A. P. Engelbrecht, "A study of particle swarm optimization particle trajectories," *Inf. Sci.*, vol. 176, no. 8, pp. 937–971, Apr. 2006.

[26] Z. H. Zhan, J. Zhang, Y. Li, and H. S. H. Chung, "Adaptive Particle Swarm Optimization," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 39, no. 6, pp. 1362–1381, Dec. 2009.

[27] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments-a survey," *IEEE Trans. Evol. Comput.*, vol. 9, no. 3, pp. 303–317, Jun. 2005.

[28] J. Liu, H. Ma, and X. Ren, "Tracking Multiple Targets with Adaptive Swarm Optimization," in *Applications of Evolutionary Computation*, 2011, pp. 194–203.

[29] J.-R. Zhang, J. Zhang, T.-M. Lok, and M. R. Lyu, "A Hybrid Particle Swarm Optimization-back-propagation Algorithm for Feedforward Neural Network Training," *Appl Math Comput*, vol. 185, no. 2, pp. 1026–1037, Feb. 2007.