Available online at: https://ijact.in

Date of Submission	03/09/2019
Date of Acceptance	26/11/2019
Date of Publication	03/12/2019
Page numbers	3494-3500 (7 Pages)

Cite This Paper: Nisha V., Vimala S., Hierarchy based least square approximation and interpolation method for resource allocation in cloud environment, 8(11), COMPUSOFT, An International Journal of Advanced Computer Technology. PP. 3494-3500.

ISSN:2320-0790

An International Journal of Advanced Computer Technology

HIERARCHY BASED LEAST SQUARE APPROXIMATION AND INTERPOLATION METHOD FOR RESOURCE ALLOCATION IN CLOUD ENVIRONMENT

Nisha V[1], Dr. Vimala S[2]

[1]Research Scholar, Mother Teresa Women's University,
[2]Associate Professor, Mother Teresa Women's University,
Department of Computer Science, Kodaikanal,
Tamil Nadu, India.

Abstract: Cloud resource allocation is a complex process because the Service Level Agreement(SLA) parameters and Quality of Service (QoS) parameter should be satisfied before allocating resources to the user. In this paper, resources are allocated to the user based on user requested memory and bandwidth. Weighted Least Square Approximation method is used to match the user required memory and bandwidth with the available resource memory and bandwidth. Weighted Least Square Approximation method produces a set of line equations that match the user request with the resource. For each parameter set of the line equation will be produced. Iterative Interpolation Technique is used to predict suitable resources by using the set of line equations. Data are tested in CloudSim, the proposed method provides an optimal solution in all test cases.

Keywords: Cloud computing; resource allocation; least square approximation; processing elements; optimal solution.

I. INTRODUCTION

Cloud computing is a hasting technology in the field of distributed computing. Cloud computing can be used in applications that involve storing data, data analytics and IoT applications [1]. Services provided by Cloud computing are IaaS (Infrastructure as a Service), PaaS(Platform as a Service), and SaaS (Software as a Service) [2]. Through the internet, cloud users can submit their request for each service to the service provider. The service provider is responsible for managing the resources to fulfill the requests generated by users. Service Providers are responsible for scheduling the incoming requests (tasks) and to manage their computing resources efficiently. Task scheduling and resource management permits providers to maximize revenue and the utilization of resources up to their limits. Task scheduling is the process of arranging incoming requests (tasks) in a certain order. If tasks or resources are not scheduled properly it may lead to longer waiting time, turnaround time, response time and completion time.

The scheduling process allocates the tasks within the workflow on suitable resources based on certain scheduling criteria such as execution time, cost, budget, reliability and deadline. Basically these scheduling objectives are categorized into two perspectives: Cloud Service Provider and Cloud User.

Cloud Service User Objectives are:

• Makespan: Overall time required to execute.

• Cost: The service cost of the user paid to the Service Provider.

• Deadline: Deliver of the services, within a stipulated time.

• Reliability: The extent to which the job is executed by ensuring the availability of resources and network connection.

Service Provider objectives are

• Load balancing: Load balancing means the scheduler distributes the workload to the available resource evenly. Balancing of load over the resources also improves resource utilization and overall performance.

• Resource utilization: Available resources must be utilized efficiently.

• Rescheduling: Rescheduling is an overhead to the scheduling process because it leads to re-evaluating the schedule and the cost of data movement is high. Rescheduling is essential due to the following reasons such as processor being heavily loaded, shutdown of Virtual Machines, System error.

• Energy efficiency: Energy consumption will be high when CPU's are not properly utilized and leads to produce more heat, which may harm the surrounding environment.

Proper allocation of resources leads to reduction in energy consumption, rescheduling overhead, Makespan and their associated cost, and also balancing of workload among servers. This work proposes a multi-objective resource distribution model in which a level of scheduling is associated with hierarchy.

The problem decomposed into three layers. First layer matches user requested memory with resource memory using Least Square Method. Second layer matches user requested bandwidth with resource bandwidth. Third layer finds minimum average completion time by using iterative interpolation method.

The task and resources are balanced by weighted least square approximation (WLSA) with minimum of error. A line equation is obtained from each type of parameters. The iterative interpolation method used to predict the resources by hierarchy order.

The rest of the paper is organized as follows. Section II presents related works Section III describes the proposed method and algorithm. Proposed algorithm is tested by data set in Cloudsim environment in section IV. The results are tabulated and discussed. Finally, Section V concludes this paper.

II. RELATED WORK

BaominXu et.al [1] present an approach of dual fairness constrains. The first constraint is to classify the users according to the QoS parameters such as bandwidth and completion time. The second constraint is on allocating resources according to the QoS preference.

LiyunZuo et.al [2] proposes a framework that includes task manager, resource manager and scheduler. The task manager manages task requests and classifies the tasks. The resource manager is used to collect the information of resource loads and sort resources according to their loads. The scheduler allocates the tasks to resources. The interlacing peak scheduling method has been incorporated as part of the scheduler.

Narander Kumar et.al [3] proposes a two-phase resource allocation technique. In the first phase, an open market-driven auction process is followed by a preference is driven payment process. In the second phase, service provider will find the average bit value at a particular time and it selects the user when the bit price is greater than the average bit rate that the user will be selected.

MehiarDabbagh et.al [4] illustrates the number of virtual machines (VM) requests, to be arriving at cloud data centers. The method estimates the number of physical machines (PM) using weighted least square method. All parameters are assigned in a single equation.

JiLia et.al [5] is addressed the allocation optimization challenge in the context of the Greedy-Based Algorithm. In order to prove our opinions the following steps are taken: classify tasks based on QoS and to select the appropriate branch of the function. The process of resource provisioning is conducted by Service Level Agreements (SLA) and parallel processing. Recent works allocate the resources by a single SLA parameter.

Hang Liu et.al [6] introduces the robotic systems combined with cloud computing capability. The robots send requests asking for computing services and efficiently allocate its computing resources by reinforcement learning (RL). The concept of RL allows the system of load computing-intensive tasks through online learning and reduces human participation.

Anusha Bamini et.al [7] introduces the method of bacterial foraging optimization algorithm with a genetic algorithm (GABFO) for scheduling problems in cloud workflow. The concept of reliability, availability and robustness are included. The work was compared with PSO, Genetic and Genetic with BFO. The final result shows better performance and maximum resource utilization.

Zhang Q et.al [8] presents the method of estimation of resources by non-negative least square estimation approach. Least square method regression technique for estimating resource demands is presented by authors Kalbasi et.al [9]. A component-based model for resource estimate using a weighted least square estimation method is introduced by Liu Y [10] et.al.

III. PROPOSED METHOD AND ALGORITHM

The Least Square method is a form of mathematical regression analysis that finds the line of best fit for a dataset, providing a visual demonstration of the relationship between the data points by minimizing the sum of squares. Interpolation [11] is defined as the process of predicting the value of the dependent variable y corresponding to a particular value of the independent variable x when the function y is described by a set of discrete data.

A. Proposed Method

Non-Linear regression is called as error it produces uneven line with curves. The points are represented by { (xi, yi) | i=1,2,…n}, then to choose a and b to minimize the following function.

$$f = \sum_{i=1}^{n} [\, y_i - (a + bx_i)]^2$$

Taking partial derivatives with respect to (a, b) and setting them equal to zero.

$$\frac{\partial f}{\partial a} = 0$$
$$\frac{\partial f}{\partial b} = 0$$

In such a way that the parameters a and b are determined by performing a least square of the data. This regression line y = a + bx is known as normal equation of least squares. The goodness of fit is determined between the line and data. For discrete *data* $(x_0; y_0); (x_1; y_1) ..... (x_N; y_N)$, weights w_i are positive float according to the relative accuracy of data points. When a discrete data point is more accurate than the other data, a larger weight is assigned. In some cases equal weights are assigned i.e., $w_i = 1$ for $i = 1$ to N.

Where,
N is the number of tasks.
x is considered as a parameter of tasks.
y is considered as a parameter of resources.
w is considered as weights (0 < w <=1.0).
The least square normal equations are

$$b\sum_{i=0}^{N} w_i + a\sum_{i=0}^{N} w_i\, x_i = \sum_{i=0}^{N} w_i y_i$$
$$b\sum_{i=0}^{N} w_i\, x_i + a\sum_{i=0}^{N} w_i\, x_i^2 = \sum_{i=0}^{N} w_i x_i y_i$$

The variable ‘a’ and ‘b’ are calculated by solving the above two equations. The tasks are rearranged by random with normal distribution.

Each processing elements are connected with high speed bandwidth connection. In layer 1, bandwidth of tasks (requested by user) and resources are considered for scheduling problem. The allocations of tasks for VMs are done by matching tasks bandwidth with the resources bandwidth. A regression line is obtained from the data. The tasks and resources are arranged by the goodness of fit value. The parameters of the tasks are rearranged by random with normal probabilistic distribution. The error value is calculated. The minimum of error value is a criterion for optimum solution. The iteration process will be continued till maximum iteration count.

In layer 2, memory requested by the user and resources are considered for scheduling problem. The allocation of tasks for VM is done by matching user requested memory with the resources memory. The weights are taken into account to obtain a minimum error. The Tasks and resources are scheduled by the weighted least square normal equations. The parameter of tasks is rearranged randomly by each iteration and then finding the minimum of error values. Finally, the output of the layer has a set of solutions with minimum error.

The average completion time of tasks is taken in layer 3. The tasks are allocated for available VMs. Thus the output is prepared. The lines obtained from the above layers are interpolated by hierarchy. At a time any two layer's outputs are used to predict the suitable resource by iterative interpolation. Tasks are taken and iteratively balanced with available resources. Among the list of solutions, minimum solution is stored. Then the process is repeated up to reach maximum iteration count.

B. Resource Allocation Algorithm

The step by step procedure for the proposed algorithm is presented in this section. Fig.1 shows the proposed method flow diagram.

Step 1: Initialize the parameters of memory, band width requested by the users. Initialize the parameters related with resources for above cases.

Step 2: Layer 1 process started. Band width parameters of tasks and resources are taken into account. Weights are assigned from the value. (0< w < 1.0).

Step 3: Obtain the regression line from weighted least square method. User request for task and resources are compared by goodness of fit with minimum of error.

Step 4: If the error value is less, go to next step. Otherwise re-shift the resources by random with normal distribution and go to step 2.

Step 5: From the goodness of fit value tasks and resources are arranged.

Step 6: Store the result. If the total number of results more than the limit (default N x N times, Where N is the number of resources), go to next step. Otherwise re-shift the weights of resources by random with normal distribution. Go to step 2.

Step 7: Layer 2 process started. Memory of task and resources are considered. Weights are assigned from the value.(0 < w < 1.0).

Step 8: Task and resources are compared by goodness of fit with minimum of error.

Step 9: If the error value is minimum, go to next step. Otherwise, re-shift the parameters of resources by random with normal distribution and go to step 7.

Step 10: From the goodness of t value task and resources are arranged.

Step 11: Store the result. If the total number of results more than the limit (default N x N times, Where N is the number of resources), go to next step. Otherwise re-shift the parameters of resources by random with normal distribution. Go to step 7.

Step 12: Layer 3 process started. Completion time parameter is taken into account. Weights are assigned from the value. (0 < w< 1.0). Take layer 1 and layer2 output as input for layer 3 to allocate the resource to the task.

Step13: Obtain the equations from above two layers. Compute normal equations for two layer's equations (hierarchy order) using linear interpolation method. Determine the slope of the normal equation m.

$$\begin{bmatrix} P(x) & x & 1 \\ f(x_o) & x_0 & 1 \\ f(x_1) & x_1 & 1 \end{bmatrix} = 0$$

Step 14: Compute the estimated resources using iterative interpolation method. For each discrete data, slope (m_1) of normal equation is compared with slope m.

$$P(x) = \frac{1}{(x - x_0)} \left[(x_1 - x) f(x_0) - (x_0 - x) f(x_1) \right]$$

$$P(x) = \frac{1}{(x - x_0)} \begin{bmatrix} I_0(x)\ x_0 - x \\ I_1(x)\ x_1 - x \end{bmatrix}$$

Where,

$$I_0(x) = f(x_0)$$

$$I_1(x) = f(x_1)$$

Error $\varepsilon > m - m_1$

Step 15: Print the optimum result.

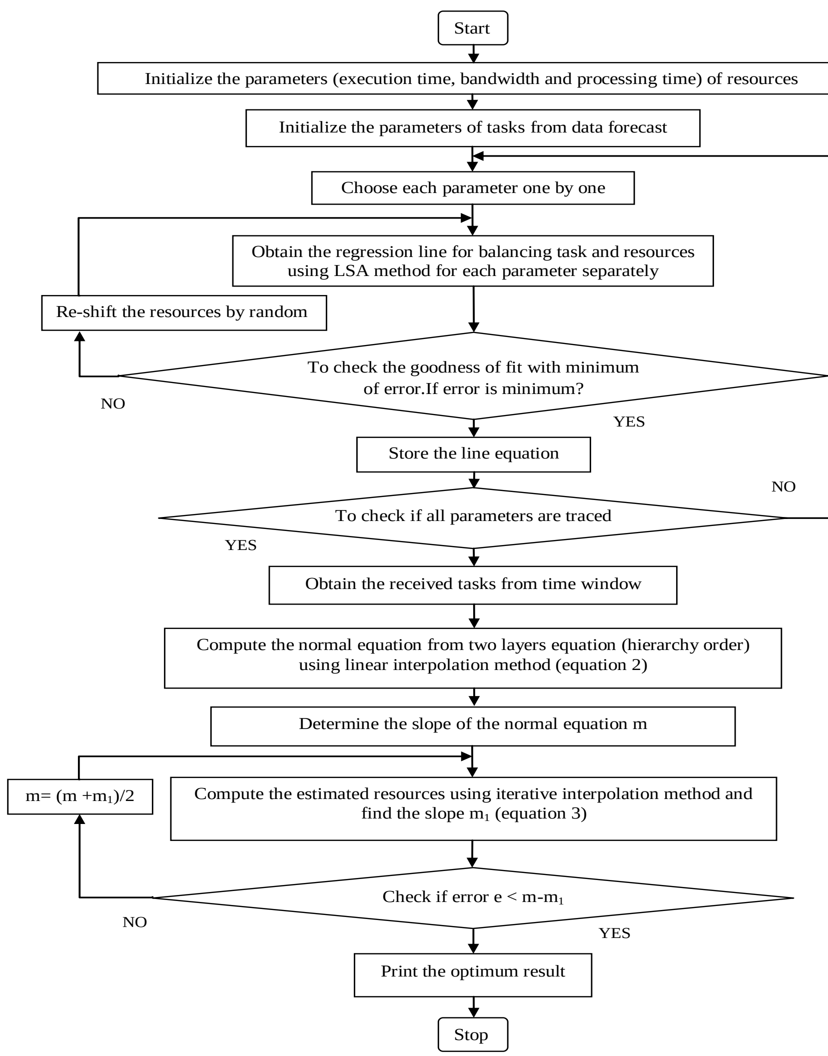

Fig 1: Flow Diagram of Proposed Method

Table 1: Results of Proposed method

Task Number	Proposed Method			Greedy Method		
	Completion Time	**Waiting Time**	**Turn Around Time**	**Completion Time**	**Waiting Time**	**Turn Around Time**
1	176.47	0	176.47	214.29	0	214.29
2	235.29	176.47	411.76	235.29	0	235.29
3	642.86	411.76	1054.62	692.31	892.31	1584.62
4	357.14	1054.62	1411.76	925.93	0	925.93
5	769.23	1411.76	2181	1851.85	2703.7	4555.56
6	882.35	2181	3063.35	430.77	0	430.77
7	1000	3063.35	4063.35	2407.41	4555.56	6962.96
8	851.85	4063.35	4915.2	270.59	235.29	505.88
9	828.57	4915.2	5743.77	892.31	1584.62	2476.92
10	492.31	5743.77	6236.08	744.19	465.12	1209.3
11	465.12	0	465.12	465.12	0	465.12
12	1185.19	465.12	1650.3	882.35	776.47	1658.82
13	1511.63	1650.3	3161.93	928.57	871.43	1800
14	328.57	0	328.57	270.59	505.88	776.47
15	611.76	328.57	940.34	1209.3	2046.51	3255.81
16	651.16	940.34	1591.5	1777.78	925.93	2703.7
17	740.74	0	740.74	285.71	214.29	500
18	461.54	0	461.54	461.54	430.77	892.31
19	514.29	461.54	975.82	837.21	1209.3	2046.51
20	400	975.82	1375.82	371.43	500	871.43

IV. RESULTS AND DISCUSSION

This section presents the results obtained from the comparative experimental study. The proposed method has implemented using CloudSim. This software platform provides tools for the design of resource allocation problem. Table I shows the results of Proposed Method and Greedy method in Cloudsim environment.

Least Square is a statistical regression method used to determine best fit line for set of points. First take the user required bandwidth and available resources bandwidth plot the points and find the best fit line and save the line equation with minimum error. Reshuffle the resources random with normal distribution and find the minimum error line equation for n times. Similarly take the user required memory and available resources memory plot the points and find the best fit line and save the line equation with minimum error. Reshuffle the resources random with normal distribution and find the minimum error line equation for n times. Then match the saved bandwidth line equation with the saved memory line equation using iterative interpolation method. According to the match allocate the resources to the task. The Proposed method is very keen to provide user required memory and the user required bandwidth. It satisfies both cloud user and cloud provider. Cloud users are satisfied because the user can get required resources with less response time, completion time, waiting time and turnaround time.

Table 2: Parameter of Dataset

Type	Parameter	Values
Server / Data Center	Number of data center	1
	Number of PE per data center	5
Processing elements (PE)	Execution Speed of PE	1000- 200000 MIPS
	Bandwidth	0.28- 3.45 kW
	Instruction Volume	128, 256,1024 (1MB), 2048 (2MB)
Task	Total No. of Tasks	20
	Length of Tasks	5000-15000 Million Instruction

The Cloudsim has the ability to create datacenters, virtual machines and physical machines and configure system brokers, system storage. In the cloud computing resource allocation, VM linked together to executive the corresponding tasks with different preferences and requirements. Experiments have been performed on a Cloudsim environment with 20 tasks with 5 processing elements.

Table 2 summarizes the parameters of the data set. Table 3 shows the summary of the Virtual machine speed. Each task and its number of instructions are presented in Table 4.

The inputs are number of instructions, required bandwidth and required memory of VM. The outputs are analyzed in terms of completion time, waiting time and turn-around time. The iterations have been repeated with different seeds for the random number that is sufficient to obtain a narrow confidence interval.

Fig 2 shows the efficient scheduling of both general tasks and resources using the Gantt chart. The chart is drawn from the time v/s virtual machine.

The tasks T14,T4,T19,T3 and T9 are scheduled to VM1. The tasks T18,T10,T5 and T7 are assigned to VM2. The tasks T11,T16 and T13 are scheduled to VM3.The tasks T1, T2, T20,T15 and T6 are assigned to VM4. The tasks T17, T8 and T12 are assigned for VM5

Table 3: VM Specifications

VM	VM Speed- No. of instruction per Millisecond	VM Bandwidth	VM Memory
VM1	70	7000	8000
VM2	65	4000	5000
VM3	43	3700	4000
VM4	85	6000	7000
VM5	27	2200	3000

Table 4: User Tasks

Task Number	No. Of Instructions	Task Number	No. Of Instructions
1	15000	11	20000
2	20000	12	32000
3	45000	13	65000
4	25000	14	23000
5	50000	15	52000
6	75000	16	28000
7	65000	17	20000
8	23000	18	30000
9	58000	19	36000
10	32000	20	26000

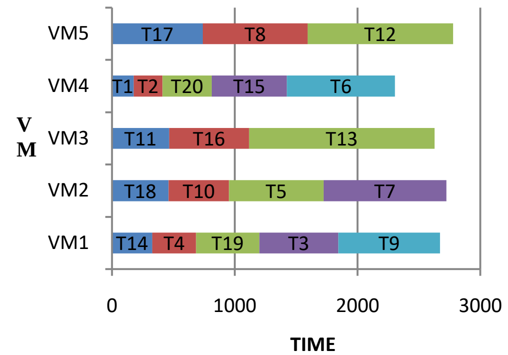

Fig 2: Time sequence diagram for proposed method.

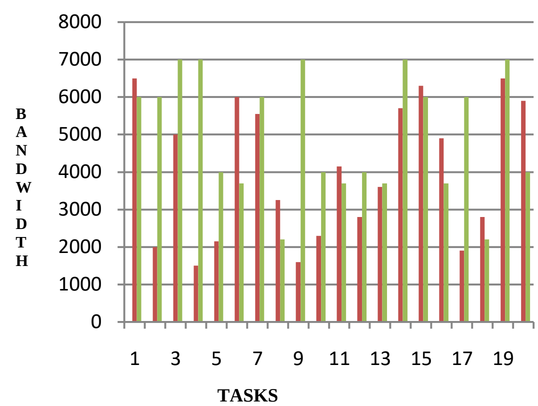

Fig 3: Bandwidth comparison of requested and allotted bandwidth using proposed method

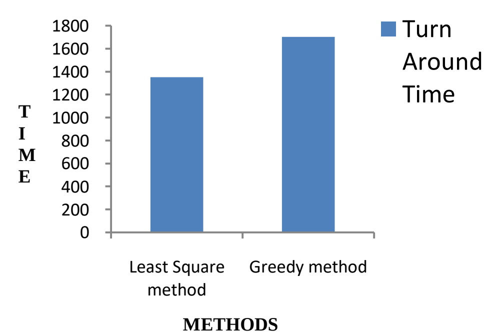

Fig 4: Turn Around Time

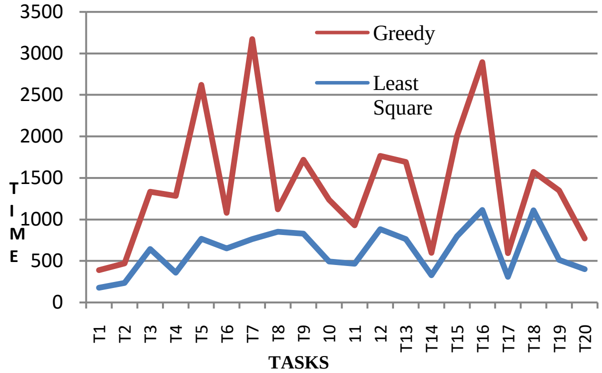

Figure 5: Completion Time

Figure three shows the bandwidth of allotted resources with the required bandwidth of the user. It is noted that the proposed algorithm and greedy algorithm presents a similar trends and comparable results regarding the utilization of resources. Such goodness of fit mechanism triggers the solution for the purpose of reallocating virtualized resources, according to the specification of tasks. Tasks would have less turn-around time in the queue as shown in Figure 4.

While comparing with an average completion time of the process, the proposed method shows good performance. However, it is important to observe two specific aspects of the proposed method. They are (i) it can take a minimum turn-around time and (ii) the capability of incorporating a bandwidth.

Figure 5 illustrates the strict consistency requirement for task completion time with optimal scheduling of computing resources. The tasks are designed and inserted into the task queue in the Cloudsim environment. The general tasks would have less waiting time in the queue.

V. CONCLUSION

Thus, the Least Square resource allocation algorithm satisfies the multi-objective proposed at the beginning of the work. Moreover, the Least square technique and iterative interpolation technique have shown the best performance while complying with the problem constraints. Mapping user required memory and bandwidth with the parameters of resource using the Least Square method gives minimum turnaround Time, completion time and allot the resources to the cloud user according to their need compared to Greedy Method. This paper considers two parameters (Memory and Bandwidth) to allocate resources to the user. By increasing the number of layers, the user can increase the parameters required and the proposed method produces the best results. In the future, Quality of Service parameters and Service Level Agreement parameters are considered to allocate resources such as response time, number of VMs needed.

VI. REFERENCES

[1] BaominXu ,Chunyan Zhao, Enzhao Hu and Bin Hu, “Job scheduling algorithm based on Berger model in cloud environment”, Advances in Engineering Software, Elsevier, Vol. 42 (2011) , pp. 419–425.

[2] LiyunZuo, Shoubin Dong, Lei Shu, Chunsheng Zhu and Guangjie Han, “A Multi-queue Interlacing Peak Scheduling Method Based on Tasks”, Classification in Cloud Computing’, IEEE Systems Journal, 2016.

[3] Narander Kumar and Swati Saxena, “A Preference-based Resource Allocation In Cloud Computing Systems”, 3rd International Conference on Recent Trends in Computing, Procedia Computer Science, Elsevier, Vol.57 (2015), pp. 104 – 111.

[4] MehiarDabbagh, Bechir Hamdaoui, Mohsen Guizani and AmmarRayes, “Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers”, IEEE Transactions on Network and Service Management, Vol. 12 (2015), No. 3, pp. 377-391.

[5] JiLia, LonghuaFenga and Shenglong Fang, “An Greedy-Based Job Scheduling Algorithm in Cloud Computing”, Journal of Software, Academy Publisher, Vol. 9, No. 4, April 2014, pp. 921-925.

[6] Hang Liu, Shiwen Liu and KanZheng, ‘A Reinforcement Learning-Based Resource Allocation Scheme for Cloud Robotics’, IEEE Access, Vol.6, 2018.

[7] AnushaBamini Antony Muthu1 and Sharmini Enoch, “Optimized Scheduling and Resource Allocation Using Evolutionary Algorithms in Cloud Environment”, International Journal of Intelligent Engineering and Systems, Vol.10, No.5, 2017.

[8] Zhang Q, Cherkasova L, Smirni E (2007) A regression-based analyticmodel for dynamic resource provisioning of multi-tier applications.In: Proc. of the 4th ICAC Conference, Jacksonville, Florida, USA,pp. 27–27.

[9] Kalbasi A, Krishnamurthy D, Rolia J, Richter M (2011) “MODE: Mix drivenon-line resource demand estimation”. in Proceedings of the 7th International Conference on Network and Services Management. International Federation for Information Processing, pp 1–9.

[10] Liu Y, Gorton I, Fekete A (2005) “Design-level performance prediction of component-based applications”. IEEE Trans SoftwEng 31(11):928–941.

[11] S.S. Sastry, “Introductory Methods of Numerical Analysis”, 5th Edition, PHI Learning Private Limited, Delhi, India, 2012, pp. 101-125.