Educational Administration: Theory and Practice
2024, 30(5), 9657-9666
ISSN: 2148-2403
https://kuey.net/
Research Article

Clustering Based Prediction For VM Workload In Green Computation

Dr. Shailesh Saxena[1], Dr. Mohammad Zubair Khan[2], Dr. Ravendra Singh[3], Mr. Ashish Agarwal[4], Ms. Priyanka Pramanik[5]

[1]Associate Professor, Deptt. of CSE, SRMS CET, Bareilly, shaileshgla@gmail.com
[2]Professor, Deptt. of CS and Information, Taibah University Medina Saudi Arabia, mkhanb@taibahu.edu.sa
[3]Professor, Deptt. of CS and IT, MJP Rohilkhand University, Bareilly, rsiet2002@gmail.com
[4]Assistant Professor, Deptt. of CSE, SRMS CET, Bareilly, agarwal.ashish01@gmail.com
[5]Research Scholar, Deptt. of CSE, SRMS CET, Bareilly, priyankapramanik20k@gmail.com

Citation: Dr. Shailesh Saxena, et al (2024), Clustering Based Prediction for VM Workload in Green Computation, *Educational Administration: Theory And Practice*, 9657-9666, DOI: 10.53555/kuey.v30i5.2585

ARTICLE INFO

ABSTRACT

In order to support the different new generation equipment and technologies, cloud computing is depends to deal with the bulky data. Because a rich amount of data is generated using these devices and processing such big data need the cloud servers which can scale the computational ability according to demand. On the other hands to perform computation we need huge power supply and cooling system that increases the power consumption and emission of harmful gases. Thus, need to achieve green computing by reducing power consumption of computational cloud. In this context, we found VM(virtual machine) workload scheduling can be a good strategy to efficiently utilize the computational resources and reducing power consumption of cloud server.
Basically, the physical machines contain a number of virtual machines (VMs). These VMs are used to deal with the workload appeared for processing. If we better utilize the resources then we can process large number of jobs in less amount of VMs. Additionally, we can also turn off the ideal machines to reduce the power consumption. In this context the proposed work is motivated to work with VM scheduling techniques to achieve green computing. In recent literature we also identified that there are two kinds of VM scheduling approaches active and proactive. The proactive technique is more effective as compared to active approaches, due to prior knowledge of the workload on VM. So, in this paper we proposed green cloud predictive model for VMs workload using unsupervised learning (i.e clustering) like K-Mean, K-Medoid, Fuzzy C-Mean (FCM),Self-Organizing Map(SOM) to predict the future workload for VM's scheduling and find the efficient clustering among them for workload prediction in view of green computing. The efficiency of clustering-based prediction is measured on parameters like accuracy, error rate.

Keywords: Virtualization, Power Consumption, Green Environment, Proactive, Clustering, Prediction.

1. Introduction

The term "Cloud" refers such type of server(s) that can access through Internet, and also access the software, storage and other resources of those servers. These servers are situated in large data centers located on the different coordinates of the world. So the cloud computing facilitates the individual users as well as IT companies to perform their computing without manage physical servers themselves. Now it is not mandatory to have application software(s) or storage on user's machines due to the computing over cloud. Users are able to store their files or data over cloud and also access their files and application software for computing to the cloud from anywhere. Google Drive, Drop box, Cloud E-mail Services (Gmail and MS Office 365) are some of examples used by users as a cloud services.
Cloud computing provides a variety of on-demand technologies and services like application software, storage and processing power on pay-as-you need basis. In place of having their own computing infrastructure or server, IT companies can access to anything (software, storage and processing) from a cloud service provider on rent basis. It became more beneficial to them in terms of reduction in the upfront cost and complexity of

purchasing and maintaining their own IT infrastructure. In cloud computing they can pay only for what they use and when they use. As per the literature the starting of cloud computing as a technology has been define around since the early 2000s, but the concept of computing-as-a-service has already been started around near about in 1960s an edge of mainframe, when companies booked timeslot for computing on mainframe, rather than to buy one themselves. After the rise of personal computer (PC) this 'time-sharing' services were turned into personal computing, and then it overtaken by the corporate data centers where companies would able to store their vast amounts of data. But due to these changes in the concept of renting computing, the concept of computation has resurfaced many times in between 1990s to 2000s as application service providers, utility computing, and grid computing. These changes in computation were followed by cloud computing, which defines the emerging development in computation as providing the three mode of computing—software as a service(SaaS), platform as a service(PaaS), infrastructure as a service(IaaS).
More than one third of worldwide IT industry use the cloud services either in terms of infrastructure, application, storage, data analysis or security, shows the popularity and efficiency of cloud computing in recent days. It is also
predicted that half of global IT industry using the cloud in the end of 2021 and all will be move on cloud upto 2023.
No doubt that users have many benefits in using the cloud computing tools, but on the other have it also have some issue like incredible consumption of energy within the range of many megawatts [5], high omission of CO_2 have negative consequences over environment and related high pricing of computing effected the cloud service providers for the operation and maintenance of cloud resources [6].
The latest development of resource-based processing technology makes it more difficult for energy consumption to reach the best level. However, since the amount of power consumed by electronic equipment(s) is mainly affected by the way it is used. So it is a key challenge in cloud computing, how effective use of resources and the use of useful resources is define. To address these challenges various energy-saving technologies based on virtualization is prescribed [7].
Cloud computing require thousands of watts of energy for providing the efficient services to customers and to executethe demands of user(s) in numerous categories, consume large amounts of energy. Virtualization, containers, and computational development, as well as machine-efficient methods are feasible hardware developments to reduce the consumption of energy [9]. We can accomplish the intensity of green cloud by maximizing the elimination of energy consumption and boosting the efficiency of cooling and electrical power in the Cloud data center infrastructure. Last year, cloud sources consumed around 1.1% to 1.5% of the world's energy. By the end of 2021, cloud data centers are estimated to consume more than 140 TH [10].
Green cloud computing is intended to reduce energy consumption by enabling for more efficient operation and utilization of computing resources.The foremost vital objective of novice cloud computing is that the management of energy-conscious data, virtualization algorithms, economic policies, resource management, employment and activities within the field, moreover as several land and eco-friendly standard technologies for upcoming years [11].***Transfer of service(s) is an effective way to manage cloud resources, especially in the case of short-term and long-term use of cloud resources. Efficient use of green cloud assets, put unused resources in power saving mode (ideal/sleep/OFF) to reduce energy consumption through active resources in the cloud[12].***In terms of analysis, due to excess resources, both the short-term and long-term use of IT infrastructure in cloud services face high operating costs[13].

2. Background and Motivation

Cloud Computing (CC) presents a multi-tenant architecture that is used by several concurrent users, each of whom behaves differently. At the server level, such heterogeneity generates a rapidly variable load and develops new usage patterns. VM interference plays an important role in load balancing between them. Forecasting server load is essential to ensure efficient use of resources
Hajer Toumi et al [1] propose, a prediction model of real-time server loads based on the classification of incoming processing and the detection of VM interference. Incoming job classification is used to capture trends in inbound workloads, and VM interference detection is intended to capture the percentage of interference. Finally, load prediction takes into account server resource usage, VM interference rates and receives load trends. They offer an advance version of HAT (Hoeffding Adaptive Tree) reinforced with ensemble drift detectors. The results show that Real-Time Server Load Prediction System can handle changes and better accuracy in minimum evaluation time and shorter memory space.
Many challenges such as VM migration, server allocation, scalability, availability and fault tolerance yet to be address for the efficiency of cloud computing. ***Kefaya S. Qaddoum et al [2]*** are concerned with the concept of servers' load balancing; a method of distributing the load among various nodes of all distributed systems in order to utilize resources and work response time to improve scalability and user satisfaction. Readjustment of load through dynamic resource allocation is presented to adapt to changing needs. The author presents an EANN (Resilient Adaptive Neural Network) with modified adaptive smoothing errors to build an evolving system that predicts VM load. To evaluate the specified method, the author used a cloud simulator to perform a series of simulations and compared it to a previously proposed approach. The results

show that the proposed method is better approach than other approaches.
Cloud systems require accurate server load prediction in order to maximize resource usage, reduce energy consumption, and ensure the QoS (quality of service). ***X. Xiaolong et al [3]***analyze the load characteristics of cloud servers with benefits and downsides of standard server load prediction algorithms and then proposed a new load prediction algorithm for servers based on CM-MC an integration of cloud models (CM) and Markov chains (MC). The algorithm leverages past data samples for training and then uses Markov prediction theory to obtain the degree of membership, based on the predictive value summary used in the cloud model. Experimental results show that the proposed prediction algorithm is more accurate than other common load prediction algorithms. CMMC-based server load prediction algorithms are suitable for cloud systems and can help to minimize the consumption of energy.
The primary choice for managing the computing resources utilization is host load prediction, and its accuracy is crucial to meet the SLA (Service Level Agreement). In comparison to grid computing, host load data in the cloud has a high level of volatility and noise, and standard approaches have a low predictive accuracy when dealing with load of host machine, Thus, ***Hengheng Shen et al [4]*** has proposed a BiLSTM (Bidirectional Long Short-Term Memory) based prediction method for host load. The BiLSTM-based method improves LSTM and LSTM Encoder-Decoder (LSTM-ED) memory capability and nonlinear modelling. The BiLSTM-based approach produces more accuracy than other earlier models.
Performance of dynamic clouds depends on the efficiency of its load balancing and resource allocation. ***Mahdee Jodayreea et al [6]*** carried out preliminary research is made on the prediction method of dynamic resource allocation. By modeling cloud services based on upcoming workload predictions, effective cloud resource management can be achieved, which is more effective than static allocation methods. Also introduce a predictive rule-based load balancing called Cicada. CloudSim simulator is used to simulate the proposed algorithms with minimum computation and faster workload balancing. The result shows the effectiveness of a workload balancing method with lower consumption of energy.
Cloud computing is becoming an alternative to promoting QoS for data-driven applications. The database management system that supports the deployment of cloud applications should be available. Manysolutions use database replication as a strategy to distribute database transactional workloads across replicas. In this way, load balancing methods improves the utilizationof resources; however, certain decisions use the current state of the database service to make decisions. ***Carlos S. S. Marinho et al [7]*** provide predictive load balancing of services for databases replicated in the cloud.
Carlos S. S. Marinho et al [8] presented a load balancing solution for replicated cloud databases that is both predictive and elastic. Experiments have shown that using prediction models to predict possible SLA violations in time series that indicate workloads of cloud-replicated databases might be beneficial.
In a multi-tenant cloud, it is usually necessary to access the performance counters on the hypervisor at the system level. ***Hamidreza Moradi et al [9]*** presented uPredict, a performance prediction framework based on user profiles for a single VM application in a multi-user cloud. Here, three micro-benchmarks were devised to assess VM's CPU, memory, and disk contention. Applications and VM specific prediction models can be derived via regression and neural networks based on the measured performance of applications and micro-benchmarks. These models can be used to predict application performance using analyzed resource conflicts. The authors tested uPredict in a private cloud and two public clouds using representative benchmarks from PARSEC, NAS Parallel Benchmarks, and CloudSuite. The results show that the average prediction error lies in between 9.8% to 17%, for the private cloud, while 4% for public cloud. There is an intelligent load balancing solution based on uPredict, which effectively minimize the execution and response time.
Cloud computing is specifically designed for this research task and is defined as a virtual infrastructure that represents shared data and communication services. Load balancing is another important aspect of weight balancing across multiple servers. It is a mechanism for dynamically and frivolously distributing workload across all servers. Animal Migration Optimization (AMO) is an algorithm that motivates animal behavior. Different animal behavior patterns are considered during the migration process. From one place to another, follow three rules: avoid collisions with neighbors, move in the same direction with neighbors, and keep a distance from neighbors. This work uses loops, an uncontrolled time, and the algorithm assigns virtual machines to tasks without taking over load information. The task length, resource size and priority increase the execution time and response time of the task. ***Sudhanshu Mittal et al [10]*** applied AMO technology to improve load balance. Initially the data center works according to population and then AMO works. The location of the animal is determined by the utilization rate of the data center. CloudSim is used to provision and visualize cloud infrastructure segments. Result of simulation shows that the proposed technology is much better than the existing technology.
One of the major challenges in networks is Load Balancing. ***Saher Manaseer et al [11]*** used static variables with the new proposed algorithm "MEMA Technique". In this algorithm few steps are added to the weighted round robin (WRR). Moreover, a comparison of performance between the (WRR) and MEMA is presented.
Xiuze Zhou et al [12] proposed a method to evaluate the load status of server nodes using AHPGD (Analytic Hierarchy Process Group Decision). We trained using HHGA (Hybrid Hierarchical Genetic Algorithm) to optimize RBFNN (Radial Basis Function Neural Network). AHPGD creates aggregate metrics

for VMs in the cloud and is an input parameter for predictive RBFNNs. The author is also combined with a weighted round robin algorithm that calculates and continuously updates the weight value for a node using the predicted periodic load value for that node, based on the HHGA-optimized AHPPGD and RBFNN. We propose a dynamic load balancing scheduling algorithm. On the other hand, it retains the strengths of the static weighted round robin algorithm and avoids its drawbacks.

Cloud computing is attracting great attention due to the growing demand from organizations that offload compute-intensive tasks to data centers. Meanwhile, data center infrastructure consists of hardware resources that consume a lot of energy and increase carbon emissions. Cloud data centers require VMs to be assigned to various PMs (physical machine) to minimize wastage of resources and increase energy efficiency. Finding the right solution is complex in large data centers. ***Usman Mohammed Joda et al [13]*** proposed an EFPA (energy-oriented flower pollination algorithm) for VM mapping. The proposed framework enables virtual machines to be allocated with a focus on minimum energy consumption. This mapping uses a strategy called DSP (Dynamic Switching Probability). The framework quickly finds the best solution for balancing local and global searches. By prioritizing the allocation of performance-based resources to multiple VMs, the processor, memory, and PM memory limitations are taken into account. Modeling is done in consideration of the workload of the earth. EFPA is significantly superior to more advanced methods in terms of power consumption, including Power-Aware Genetic Algorithm (GAPA), ordering of exchange migration (OEM) with Ant Colony optimization, and FFD (First Fit Dismissal). This means that the performance and environmental sustainability of the data center has been improved by reducing the consumption of energy as well as emissions of CO_2.

Cloud computing has become an impressive solution for solving the issues related to storing and processing large amounts of data due to its characteristics like high speed, low pricing, on-demand working and pay as per use only. While green clouds have focus on such type of computing that deals with eco-environment, Low CO_2emissions, energy efficient with maximum utilization of resources and reusability. Environment will be protected from negative impacts of cloud computing, through adopting green computing, service providers update their infrastructure of computing for eco-environment. The research related to green computing broadly focuses on the design of eco-friendly clouds, which have characteristics like virtualization, power management, workload balancing, high performance computing, green data centers, recyclability, reusability etc. ***Archana Patil et al [14]*** present a study on green clouds and their properties as part of their research on green clouds. They highlight the importance of green computing in the present trend of computing, and also focus on challenges for future research in this field. This study helps researcher(s) to learn about green cloud or green computing.

Increasing energy demand, the costs of data center(s) are becoming a concern. 'Green Data Center(s) applies to energy-conscious, energy-efficient, and CO2 emissions reducing architectures, protocols, tools, infrastructures, and algorithms. Data Center(s) are equipped for peak processing. But, most of the time, servers are seen to be idle. Idle servers and related components of the network consume a significant amount of resources. There are also ways of reducing energy wastage, the power costs. Through utilizing a green computing method, the data center(s) become green to reduce the electricity consumed by the data center(s). The facilities in the data center(s) can be constructed in such a way that they use green computing to consume only minimal energy. ***P Aishwarya Naidu et al [15]*** describe potential research enablers for the green data center and green metrics specific to data center(s) and discuss energy-saving solutions for servers, network, and other green solutions.

Effective VM management is critical for reducing energy use, improving profits, and avoiding SLA violations. VM placement can be categorize into reactive and proactive/predictive methods which attempt to enhance VM placement results by projecting future loads or resource needs using prediction methods. ***Mohammad Masdari et al [16]***proposed a study on the active methods of VM placement and classified them according to their prediction methods. Describe how each scenario applies predictive algorithms to achieve more efficient virtual machine placement with less overhead. Compare factors such as estimated parameters, simulation software, workload data, energy management methods, and predictive factors to clarify more details. Finally, it focuses on current themes, trends and areas of open research in the future.

The rapid expansion of multimedia and communication network services like VoIP (Voice over Internet Protocol) is creating a resource crisis from two perspectives: first resource redundancy with energy loss and second Lack of resources with overloading. In cloud computing scaling of resources is allowed according to its demand. Meanwhile, a new concept of Software-Defined Networking (SDN) can provide a global view of the network for the management of integrated resource. NFV, a network function-based virtualization can also be used to implement of different network devices and functions virtually. To manage the resources of virtualized cloud VoIP centers, ***Ahmadreza Montazerolghaem et al [17]*** introduce GreenVoIP, an energy-efficient architecture. This framework not only prevents overload but also supports green computing by conserving energy by regulating the amount of VoIP servers and network equipment, such as switches. Finally, Green VoIP is tested and implemented on real-world systems such as Floodlight, Open vSwitch, and Kamailio. According to the findings, the framework can reduce the number of active devices, overloading, and QoS needs.

The increasing computational work and communication traffic puts a heavy burden on cloud data centers, resulting in high consumption of energy. To reduce this burden, edge computing has been proposed for provision the cloud services according latency-sensitive applications to exploring distributed resources (base stations, etc.). Depending on the geographic distribution of devices, edge computing is an energy-efficient platform. Therefore, it is natural to integrate Internet of Energy (EI) technology into edge computing to enable customizable energy scheduling. Edge computing supported by these EIs, both the rate of generation of green energy and the demand for data processing depend on time and space. In pursuit of high energy efficiency, it is desirable to maximize the utilization of green energy and reduce brown energy consumption. This requires careful task allocation and energy scheduling to match energy supply and demand. ***Lin Gu et al [18]*** jointly considers VM migration, task assignment, and green energy scheduling to investigate energy cost minimization issues and prove NP-hardness. Heuristic algorithms are proposed to solve the complexity. Simulations show that the algorithm can effectively reduce energy consumption.

Edge Computing, QoS provisioning has recently emerged in the expansion of cloud computing to guarantee delays for particularly delay-sensitive applications. Computational workloads can be offloaded to the server to improve the quality of experience such as transfer delay and energy consumption. However, edge server resources are too scarce to respond quickly to handle the rapid change of computational requirements. Therefore, queue delays cannot be ignored in a compute-intensive environment, such as a compute environment in which the compute environment consists of IoT applications. Also, workloads can often consume more computational energy on edge servers than in the cloud. Cooperation between edge servers and the cloud is important to provide end-user QoS for green computing. The energy-efficient and delay-guaranteed task allocation in an IoT-edge-cloud computing is investigated by ***MIAN GUO et al [19]***. They propose a delay-based workload allocation that offers the best workload allocations among local, neighbors, and cloud workloads with the least amount of energy usage and the best delay assurance. A delay-based workload allocation (DBWA) based on Lyapunov drift-plus-penalty theory is used to solve the problem. Energy efficiency, and delay guarantee are all demonstrated by the proposed analysis and simulation findings.

A problem of Vector Bin Packing (VBP) for minimizing the number of PMs used to host VM placement requests in cloud data centers, FFD (First Fit Decreasing) and its variant is widely used. A new variant of FFD, FFD-AR (FFD with aggregate ranking), has been proposed by ***Saikishor Jangiti et al [21]*** for VM placement. Simulations were performed using two data sets on AmazonEC2: the triggered data set and the synthetic data set. The FFDAR results show that the packing efficiency is superior to that of the other variants. FFDAR can be applied to a wide range of applications such as production planning and logistics.

3. Proposed Predictive VM's Scheduling

In this proposed clustreing based predictive model for VM workload in green computing, we have two phases as shown in Figure 1-

Phase I - Data Clustering
1. Request Clustering
2. Request Decomposer

Phase II - Workload Prediction
A brief discussion of these phases is given in this section so that readers have a clear picture of the proposed effort. Figure 1 is refer for illustration of it throughout this section.

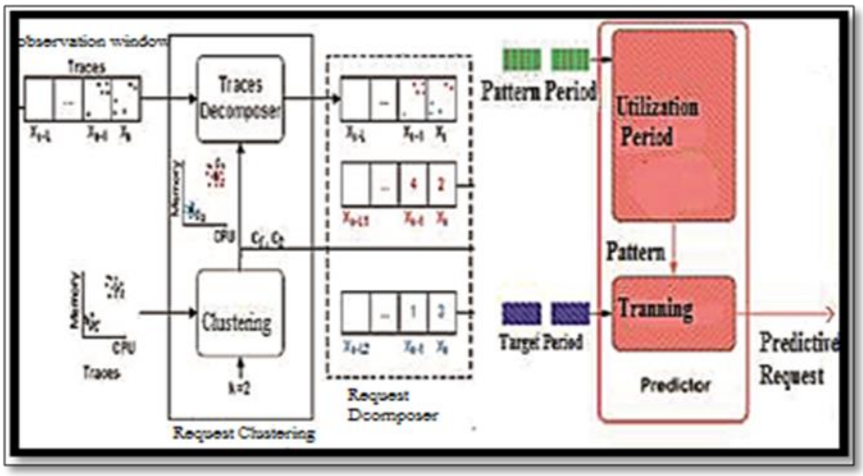

Figure 1: Clustering Based Predictive Model for VM Workload

Phase I. Request Clustering

Our proposed workload prediction approach based on clustering is depends on observing the current or past nature of workload during some particular period of time in order to predict the upcoming workload of similar type in a certain period of time of future. User's request for VM in cloud generally comprises multiple cloud resources (e.g., CPU, RAM, memory, bandwidth, etc.).

The requirenment of resources vary user by user and time by time. So this multi-resource nature of the requests with variation in demand presents unique challenges to develope a prediction techniques for upcoming user(s) request. Therefore, it is too difficult and impractical to define the prediction about the demand of each type of resource separately. To handle this issue, we categories each request for VM according to the resource type comprise by it. This is the concept of clustering an unsuperwise learning. On the basis of these categories we will able to define a prediction of user(s) requests in future. The working of phase I further divided in two steps :-

1) Request Clustering:In this step, all user(s) request is categories in form of cluster on the basis of resource demand associate with it, as each request is mapped into only one cluster. It means that all requests mapped into same cluster have requested similar type resource. Generally, N type of resources such as CPU, RAM, memory, bandwidth, etc are associated with each request for VMs in cloud. In order to divide these requests into multiple categories, we represent each request as a point in the N dimensional (RN) space. As the data set of user(s) request on Googlecluster [8], requirenment of two types of resources, CPU and RAM, are associated with each request. Thus two dimensional (R2) space is used to represent these requests shown in figure 1, where each point is a request and the required amount of CPU and RAM associated with the request are shown through the coordinates of R2space.Clustering of theserequest points into k=2 clusters is done using unsupervised learning like-k-Means, K-Medoid, FCM and SOM algorithms.

As shown in Figure 1, clustering algorithms takes requests as an input from the Google request trace and the number of clusters in terms of k, and then produce k clusters of request as outpout, each specified by its center point. In this process first we define the preprocessing of input dataset and then the clustering algorithms are employed on dataset for preparing the comparative performance study among the implemented clustering algorithms. For the case of the Google data, where requests have two types of resources (CPU and RAM), the cluster centers are points in the R2 space. It is enough to define clustering once on large set of request traces of the cloud. But if the characteristics of requests have change significantly over timethen clusteringcan be performed more than once accordinglly. In the proposed simulation, we use the Google request traces of a particular time period so that clusteringis performed only once to extract thecharacteristics of the requests submitted for VMs in the cloud.

2) Request Decomposer: After clustering of request data set, we have two types of clusters (k=2) with their centerpoints. One cluster is associated with CPU bound requests and another one is associated with RAM bound requests. These cluster of requests are given to the request decomposer (shown in Fig. 1) to define the distribution of user requests for prediction of requests. The prediction of future request is based on the estimation of request's distribution in any predetermined period. It is observe that the workload (requests) in the cloud have certain pattern of repetition according to the time frame and it repeated periodically. For example, the request pattern in weekends will be a similar and high from the working days.

In request decomposer module, we have a pattern window based on some time frame (size of window) for each cluster to define the distribution of past requests i.e the pattern of request's repetition.Here we have a two days pattern as a size of patternwindow and it is further divided into t timeintervals per day.Suppose a prediction needsto be made at time n. In this case, slot n corresponds to time interval in each day of patten window is observed.The request decomposer tracksthe number of received requests in any time slot from the pattern window with respect to all the intervals and maps each request within the slots with any one cluster as per their need of resource for workload prediction in the next day.

Phase II. Workload Prediction

Our framework relies on tracking training data to determine the predictive weights of our workloads. Large changes in workload characteristics over time require retraining of predictors according to updated weights. Therefore, these predictive models are sometimes called static predictive models that require training. It uses an adaptive prediction system to retrain the model on new training data to change the weights, improving accuracy over time and eliminating the need to store large trucks. A predictive adaptive model estimates the number of future requests based on a specific set of weights learned during the initial training period. Then the template for each class keeps track of the number of requests received in that time interval. These tests are used to change weights. The adaptive model ensures this and goes on and on. Therefore, the adaptive method uses all detected traces to evaluate the optimal weight up to the forecast period. However, the static approach uses a portion of the training data to adjust these weights. The calculation of updating the weights every minute after observing the actual workload is the predictor's overhead. However, it takes these updates into account, improving the accuracy of the forecast over time and dynamically building the model in response to the latest request transformations.

4. Experimental Result for Comparative Analysis

The aim of the proposed work is to investigate the predictive machine learning techniques in order to better schedule the cloud VM for enhancing the performance and running cost of cloud infrastructure and achieving the green computing. Therefore some unsupervised learning approaches are compared to identify the best performing clustering approach. That can deals with cloud virtualization issues and achieving the Green Computing or reducing the power consumption of cloud servers. The VM scheduling ensures the optimal VM resource utilization and reduction in power consumption of cloud servers. Thus the given experiment includes two experimental scenarios:

Comparing Different Clustering Algorithms:

Cluster analysis or clustering is an unsupervised learning task. This includes automatically detecting data groupings. Unlike supervised learning, clustering algorithms interpret only the input data to find natural groups or clusters. In the present simulation, we compared four clustering algorithms, K-Means, K-Medoid, FCM and SOM over two parameter of performance: accuracy and error rate.

In this experimental scenario a comparative performance study is carried out among different unsupervised learning algorithms. The aim is to obtain an efficient and accurate algorithm for predicting the accurate VM resource demand to better utilization of resources. The comparative outcomes of the algorithm's performance are reported in this section.

The accuracy can be explained as the measurement of algorithm correctness. That can be measured using the ratio of total correctly classified and the total patterns to be classified. That can also be represented using the following equation:

$$accuracy = \frac{total\ correctly\ classified}{total\ patterns\ to\ classify} X1 \quad (1)$$

The error rate of an algorithm demonstrates the misclassification rate of the algorithm. That can be calculated using the following equation:

$$Error\ Rate = 100 - Accuracy \quad (2)$$

Or

$$error\ rate = \frac{total\ misclassified\ samples}{total\ samples\ to\ classify} X100 \quad (3)$$

The accuracy ratio and error rate of the clustering algorithms with different data set sizes is shown in Table 1 and Table 2 respectively. The size of dataset is increased in each experiment and performance is measured in terms of percentage (%).

Table 1. Accuracy (%) of Clustering Algorithms

Dataset Size	K-Mean	K-Medoid	FCM	SOM
100	51.85	63.35	60.09	62.91
300	53.21	71.82	63.21	66.39
600	58.29	84.63	66.53	72.33
900	61.24	87.73	67.25	74.82
1100	65.57	91.86	67.44	74.22
1300	69.41	93.52	67.52	78.15
1500	70.32	95.17	69.29	79.26

Table 2. Error rate (%) of Clustering Algorithms

Dataset Size	K-Mean	K-Medoid	FCM	SOM
100	48.15	36.65	39.91	37.09
300	46.79	28.18	36.79	33.61
600	41.71	15.37	33.47	27.67
900	38.76	12.27	32.75	25.18
1100	34.43	8.14	32.56	25.78
1300	30.59	6.48	32.48	21.85
1500	29.68	4.83	30.71	20.74

The comparative chart of accuracy ratio and error rate of the implemented clustering algorithms are shown in Figure 2 and Figure 3 respectively where the calculated accuracy and error rate of the implemented clustering is notified in Y-axis and X-axis shows the different values of dataset.

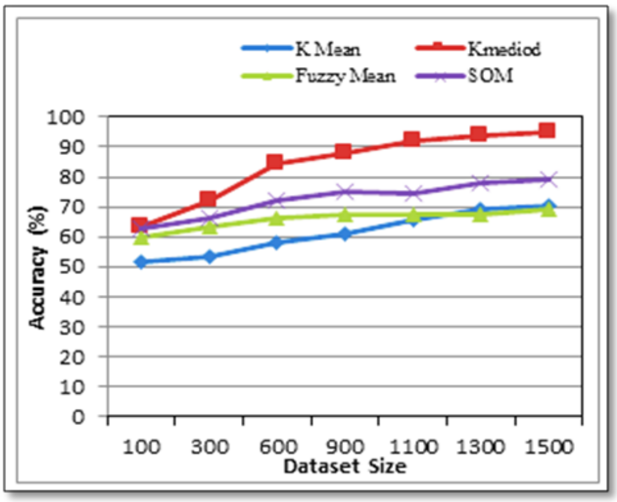

Figure 2:Comparative Chart of Clustering using Accuracy

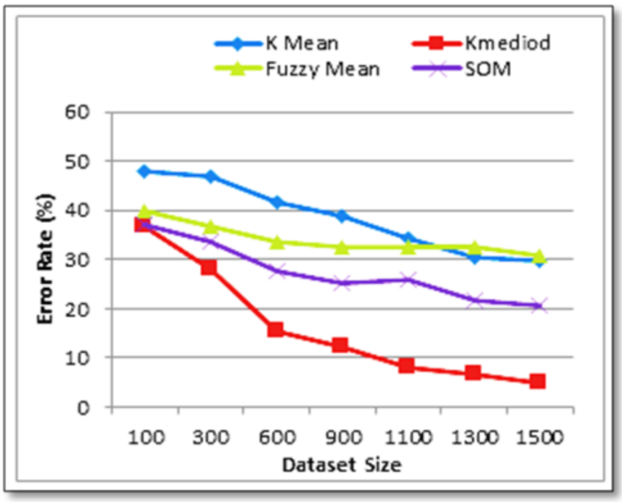

Figure 3: Comparative Chart of Clustering using Error rate

According to the obtained results shown in figure 2 and 3, the performance of the K-Medoid is found much accurate and effective, it also has lowest error rate among other implemented clustering algorithms. On the basis of these results, we say that the cluster output comes from K-Medoid is used for the effective workload prediction.

1. Conclusion and Future Aspects

We propose a green cloud predictive model for VMs workload using unsupervised learning (i.e clustering) to predict the Request workload for VM's scheduling in future. These predictions will be helpful to keep ON the right amount of needed Physical Machines in the cloud data centers. Clustering has the capacity to manage the variance of request which vary according to the need of users. For this purpose, we analyze the different clustering algorithms like K-Mean, K-Medoid, FCM (Fuzzy C-Mean), SOM (Self-Organizing Map) and find the efficient clustering among them for workload prediction in view of green computing. On the basis of accuracy and error rate, our simulation shows that the learning through K-Medoid is more effective than others to predict the future workload for VM. We use real request data set of 29 days from Google Cloud Cluster for the effectiveness of our predictive model.
In future, first we develop an energy saving framework using the proposed clustering-based workload prediction model. So that we fulfill the current demand of green computing.

Declarations:

The authors declare that they have no conflict of interest.
This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. H. Toumi, Z. Brahmi, M. M. Gammoudi, "Real time server load prediction system for the ever-changing cloud computing environment", ScienceDirect – 2019
2. K. S. Qaddoum , N. N. El Emam , M. A Abualhaj, "Elastic neural network method for load prediction in cloud computing grid", International Journal of Electrical and Computer Engineering (IJECE)- April 2019
3. X. Xiaolong, Z. Qitong, M. Yiqi, L. Xinyuan, "Server load prediction algorithm based on CM-MC for cloud systems", IEEE - Journal of Systems Engineering and Electronics Vol. 29, No. 5, October 2018.
4. H. Shen, X. Hong, "Host Load Prediction with Bi-directional Long Short-Term Memory in Cloud Computing", arXiv – 2020.
5. *P. Sudhakaran, S. Swaminathan, D. Yuvaraj, S. S. Priya, "*Load Predicting Model of Mobile Cloud Computing Based on Glowworm Swarm Optimization LSTM Network*". IJIM – 2020.*
6. *M. Jodayreea, M. Abazab, Q. Tan, "A Predictive Workload Balancing Algorithm in Cloud Services", ScienceDirect -2019.*
7. C. S. S. Marinho, E. F. Coutinho, J. S. C. Filho , L. O. Moreira, F. R. C. Sousa, J. C. Machado, "A Predictive Load Balancing Service for Cloud-Replicated Databases", IEEE – 2017
8. C. S. S. Marinho, L. O. Moreira, E. F. Coutinho, J. S. C. Filho, F. R. C. Sousa, J. C. Machado, "LABAREDA: A Predictive and Elastic Load Balancing Service for Cloud-Replicated Databases", Journal of Information and Data Management, Vol. 9, No. 1, June 2018.
9. H. Moradi, W. Wang, A. Fernandez, D. Zhu, "uPredict: A User-Level Profiler-Based Predictive Framework for Single VM Applications in Multi-Tenant Clouds", arXiv – 2019.
10. S. Mittal, Prof. M. Dubey, "AMO Based Load Balancing Approach in Cloud Computing", IOSR Journal of Computer Engineering (IOSR-JCE) – 2017.
11. S. Manaseer, M. Alzghoul, M. Mohmad, "An Advanced Algorithm for Load Balancing in Cloud Computing using MEMA Technique", International Journal of Innovative Technology and Exploring Engineering (IJITEE) January 2019.
12. X. Zhou, F. Lin, L. Yang, J. Nie, Q. Tan, W. Zeng, N. Zhang, "Load balancing prediction method of cloud storage based on analytic hierarchy process and hybrid hierarchical genetic algorithm", SpringerPlus (2016).
13. U. M. Joda, A. S. Ismail, H. Chizari, G. A. Salaam, A. M. Usman, A. Y. Gital, O. Kaiwartya, A. Aliyu, "Virtualization oriented Green Computing in Cloud Datacenter: Flower Pollination Approach", Springer – 2018.
14. A. Patil, R. Patil, "An Analysis Report on Green Cloud Computing Current Trends and Future Research Challenges". Elsevier SSRN – 2019.
15. P. A. Naidu, P. Chadha, V Nalina, "Efficient Strategies for Green Cloud Computing", Journal of Network Communications and Emerging Technologies (JNCET) Volume 10, Issue 6, June (2020).
16. M. Masdari, M. Zangakani, "Green Cloud Computing Using Proactive Virtual Machine Placement: Challenges and Issues", Springer Nature B.V. 2019.
17. A. Montazerolghaem, M. H. Yaghmaee, A. L. Garcia, "Green Cloud Multimedia Networking: NFV/SDN based Energy-efficient Resource Allocation", IEEE Transactions on Green Communications and Networking – 2017.
18. L. Gu, J. Cai, D. Zeng, Y. Zhang, H. Jin, W. Dai, "Energy efficient task allocation and energy scheduling in green energy powered edge computing", Elsevire – 2019.
19. M. Guo, L. Li, Q. Guan, "Energy-Efficient and Delay-Guaranteed Workload Allocation in IoT-Edge-Cloud Computing Systems", 2019 IEEE
20. L. Leonard, B. Miles, B. Heidari, L. Lin, A. M. Castronova, B. Minsker, J. Lee, C. Scaifed, L. E. Band, "Development of a participatory Green Infrastructure design, visualization and evaluation system in a cloud supported jupyter notebook computing environment", 2018 Elsevier Ltd.
21. S. Jangiti, E. S. Ram, V. S. S. Sriram, "Aggregated Rank in First-Fit-Decreasing for Green Cloud Computing", Springer Nature Singapore Pte Ltd. 2019.