Journal of King Saud University – Computer and Information Sciences 34 (2022) 6481–6490

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Host utilization prediction using hybrid kernel based support vector regression in cloud data centers

Priyanka Nehra, A. Nagaraju *

Central University of Rajasthan Bandarsindri, Kishangarh, Ajmer 305817, India

ARTICLE INFO

Article history:
Received 23 November 2020
Revised 26 March 2021
Accepted 25 April 2021
Available online 28 May 2021

Keywords:
Cloud computing
VM consolidation
Host utilization
Support vector regression
Prediction

ABSTRACT

The rapid growth in the cloud data center needs a dynamic resource provision to maintain the Quality of Services parameters. To guarantee it, Virtual Machine Migration as part of VM Consolidation has a significant role. Efficient VM migration requires knowledge of the host's future utilization in advance. Because of the high variation in cloud resource usage and dynamic workloads, predicting host utilization using utilization history is challenging. This paper proposes a Support Vector Regression-based methodology to predict a host's future utilization using multiple resource's utilization history. A Hybrid Kernel function that includes radial basis function and polynomial kernel function has been proposed and then trains the Support Vector Machine using multiple-resource utilization history. Compared to the existing approaches: multiple linear regression-based prediction, Euclidean distance, and Absolute Summation based regression, the proposed method performs better in terms of root mean square error, mean square error, mean absolute percentage error, mean absolute error, and R^2. The result section concludes that on evaluating error percent, the prediction error is 16% for the proposed approach and predicts host utilization with 7%, 64%, and 67% more accuracy than MRHOD, MDRHU-ED, MDRHU-AS approaches, respectively.

1. Introduction

Cloud computing is a distributed and parallel system that comprises interconnected homogeneous and heterogeneous physical machines, offering on-demand services based on the pay-as-usage model (Hu et al., 2011; Armbrust et al., 2010). Cloud Computing has molded the way existing IT infrastructure, application, and system software are used by the customers and emerging rapidly (Buyya et al., 2010). Various govt organizations, industries, e-commerce, healthcare, and academics have emerged their services on Cloud Data Centers (CDC). Cloud data centers consume an immense amount of energy, leading to high operational cost. The data centers consume high energy consumption not just because of an enormous number of resources but also because of inefficient utilization of resources (Nehra and Nagaraju, 2019, 2019.). One solution to reduce energy consumption is to improve the utilization of resources that can be achieved with the help of virtualization technology (Jennings and Stadler, 2015; Ferdaus and Murshed, 2014). Virtualization technology allows the sharing of a physical system over multiple virtual machines. All Virtual machines (VM) behave as a virtual computer system with their CPU, Memory, Bandwidth, and other resources. Efficiently utilizing the resources of Data Center (DC) such as CPU Cycles, Memory and Bandwidth, reduce the DC's energy consumption drastically (Abdelsamea et al., 2014). Though the energy consumption needs to reduce, it must also provide services without violating Service Level Agreement (SLA).

For static workload, several scheduling mechanisms (Jennings and Stadler, 2015; Nehra and Nagaraju, 2019; Mahrishi and Nagaraju, 2012) are proposed in the literature, but nowadays, it is also essential to handle the dynamic workload. VM Consolidation has been considered a prominent approach that can handle the dynamic workload and achieve other parameters. VM Consolidation is the process to migrate running VMs from one physical machine to another physical machine (Khan et al., 2017). Migration decision depends on underloaded and overloaded host detection. To detect an overloaded and underloaded host cloud's resource management system demands an approach that predicts future resource usage to perform VM Consolidation constructively

* Corresponding author. NH-8, Central University of Rajasthan, Department of Computer Science, Bandarsindri, Kishangarh, Ajmer, Rajasthan, 305817, India.

E-mail address: 2015phdcs02@curaj.ac.in (A. Nagaraju).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2021.04.011
1319-1578/

(Amiri and Mohammad-Khanli, 2017; ; Dinesh Kumar and Umamaheswari, 2012). In the literature number of regression approaches such as linear regression, multiple linear regression has been introduced for prediction. Yet, linear regression works only when a linear relationship exists between data.

In contrast, Support vector regression (SVR) is well known and famous for classification problems. SVR accepts non-linearity in the data, gives the flexibility to admit error, and provides a proficient prediction model to obtain the desired accuracy. Here, we have mentioned some challenges to be contemplated (Abdelsamea et al., 2014) for VM consolidation:

1. Overloaded host detection: Decide parameters to detect an overloaded host, and on detection, VM migration performed to ensure SLA.
2. Underloaded host detection: Decide parameters to detect an underloaded host; on detection, all VMs migrated, and the host switched to sleep mode to improves resource utilization.
3. VM selection: On detection of an overloaded host, a selection process takes place to select VM for migration.
4. VM placement: Placement process that identifies a suitable PM to place the selected VM.

This paper considers the challenges mentioned above and proposes a Hybrid Kernel-based Support Vector Regression method build on multiple resource usage to predict host's utilization. VM placement performed using our earlier research work Load-Balanced multi-dimensional Bin-Packing (LBMBP) heuristic to balance load among hosts.

1.1. Motivation

With the advent of cloud computing, cloud data centers (CDCs) containing computing and storage devices are growing in the range of ten to thousands. This state demands a practical resource scheduling approach and policies to handle scalability or dynamic resource provisioning. This work's primary motivation is dynamic resource management with resource usage prediction that can achieve multiple objectives. Predicting resource usage has a direct effect on over-provisioning and under-provisioning problems. Prediction can help in preventing overutilization from avoiding service level agreement violation. By preventing underutilization, can reduce the energy consumption of resources (Masdari and Khoshnevis, 2019).

The remaining paper is organized as follows: Section (2) consists of host detection techniques. Section (3) presents the proposed Hybrid Kernel Support Vector Regression (HKSVR) for predicting host utilization and algorithm for overloaded host detection. Section (4) describes implementation details and a comparison of different predicting approaches. The last section consists of the conclusion of this paper and future directions.

1.2. Contribution

- The paper introduces a Hybrid kernel support vector regression-based prediction methodology under consideration of multiple resource usage history to detect overloaded hosts in the cloud environment.
- The proposed work simulate in MATLAB and evaluated using the dataset released by Alibaba in August 2017 (A. Group, 2017).
- The proposed work compared to another existing prediction approach on different performance parameters such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), Mean Square Error (MSE) and R^2.

2. Literature survey

Dynamic resource scheduling is a crucial issue in Cloud Computing to provide service with minimum SLA violations. Researchers from the research objectives have done various prediction analyses for efficient VM consolidation in the literature. Much research has been done to predict host utilization for the detection of overloaded and underloaded hosts. Table 1 shows the comparison of various techniques from literature, their strength and weakness. Techniques that are identifying an overloaded and underloaded host categorized into two ways: static and dynamic threshold.

Static threshold reflects a fixed utilization threshold value that decides overloaded and underloaded host even for resource's dynamic nature. In dynamic threshold techniques, the threshold value varies over time depending on resource usage. Variation in dynamic threshold depends on resource usage and threshold value and decided by either of the ways: machine learning or statistical analysis:

Gahlawat and Sharma (2016), proposed an approach using support vector machine (SVM) to detect overloaded hosts in the cloud. Authors compared the proposed SVM approach to Dynamic Least Square SVM (DLS-SVM) and Least Square SVM (LS-SVM) and found that DLS-SVM predict with more accuracy and minimum error as compared to the other two.

Abdullah et al. (2020), proposed a support vector regression technique to forecast future resource usage. The proposed work uses radial base kernel function and sequential minimal optimization algorithm for training and regression. The proposed approach enhances the prediction accuracy and reduces the error percentage.

Beloglazov and Buyya (2012), proposed an adaptive threshold-based and regression-based heuristic for overloaded host detection. In the first approach, median absolute deviation (MAD), a measure of statistical dispersion, determines the threshold value MAD, i.e., the median of the absolute value of the deviation of data from its median. Deciding the threshold using MAD as compared to variance and the standard deviation is more robust. In the second approach Interquartile Range (IQR), statistical dispersion uses utilization values and the difference between the first and third quartile calculated to identify IQR. Again, the threshold value is computed using IQR. However, mean and standard deviation are prone to change and less efficient for dynamic resource utilization. Regression-based heuristic Local Regression (LR) and Local Robust Regression (LRR) models are better for detecting overloaded hosts.

Sharma and Saini (2016), proposed a dynamic VM consolidation method that meets the trade off between Service Level Agreement and energy efficiency. The authors proposed a median-based approach for auto-adjustment of upper and lower threshold values. Host machines are separated into two groups, and the median of these groups helps determine the lower and upper threshold values. Though the proposed approach minimized SLA violation or provided a maximum quality of service (QoS), the energy consumption is still higher.

Li et al. (2019), to accomplished energy efficiency and quality of service, proposed an algorithm based on discrete differential evolution. The probability distribution function is calculated for each resource, and the corresponding host machine's threshold value is evaluated. Though the model improves the Quality of Service by avoiding unnecessary overload conditions but does not focus on minimizing VM migration.

Mahdhi and Mezni (2018), use the Kernel Density Estimation (KDE) technique to predict future migration traffic and resources. The proposed work's main objective is to reduce the number of migrations by performing two-level predictions. At the first level,

Table 1
Comparison of different existing approaches.

Paper	Parameters considered	Technique used	Strength	Weakness
Gahlawat and Sharma (2016)	CPU usage	Support Vector Machine	-DLS-SVM forecast better than LS-SVM and SVM	-Compared for error metrics MAPE and maximal error only
Abdullah et al. (2020)	CPU usage	Support vector regression technique and Sequential minimal optimization algorithm	-Accuracy enhanced	-Compared with single-resource based prediction approaches
Beloglazov and Buyya (2012)	CPU usage	Adaptive (IQR and MAD) and regression (LR) based approach	-Efficient -Reduces energy consumption	-Mean and standard deviation influenced by terminal value resulting, not suitable for dynamic resource requirement
Sharma and Saini (2016)	CPU usage	Median absolute deviation approach	-Provide Service level agreement -Minimum performance degradation	-Power consumption is higher
Li et al. (2019)	CPU usage	Discrete differential approach	-Reduce energy consumption	-Not focuses on resource utilization resulting in unnecessary overhead for dynamic resource requirement
Mahdhi and Mezni (2018)	CPU, RAM and storage usage	Kernel Density estimation technique	-Reduces energy consumption	- High communication traffic
Li et al. (2016)	CPU usage	Linear Regression	-Reduces energy consumption - Reduce SLA violation	-Suitable only when linear relationship exists between data
Abdelsamea et al. (2020)	CPU, RAM and bandwidth usage	Hybrid Linear Regression	-Reduces energy consumption -Reduces SLA violation	-Prediction value normalized with a fixed point due to which performance degrades
El-Moursy et al. (2012)	CPU, RAM and bandwidth usage	Regression approach along with Euclidean distance and absolute summation	-Reduce energy consumption -Improvement in providing SLA -Reduction in the number of migration.	-Compared with Local regression technique only

the KDE technique predicts the next requested resources for each physical machine. At the second level, the capacity of the server to place a new VM is predicted. It is one of the approaches that consider CPU, RAM, and Storage for VM consolidation.

Li et al. (2016), proposed a linear regression prediction model based energy-efficient VM consolidation technique to avoid SLA violation. An approach named RobustSLR was proposed to predict overloaded hosts using utilization history and IQR approach to detect the underloaded host. The proposed model minimizes the power consumption and SLA violation. However, the consolidation process includes only CPU utilization; no other resource is considered.

Abdelsamea et al. (2020), to enhance VM consolidation proposed a multiple regression-based overload host detection algorithm, i.e., MRHOD. Numerous factors such as CPU, memory, and network are combined to develop host utilization to make it suitable for the cloud's dynamic nature. Multiple regression is applied to develop regression coefficients that help to predict utilization. Although the proposed algorithm outperforms existing approaches in terms of energy consumption, it gives poor results for SLA violation.

El-Moursy et al. (2012), considered multiple factors: CPU, Memory, and Bandwidth utilization for overloaded host detection. A multi-dimensional regression-based host detection algorithm has proposed combining CPU, Memory, and Bandwidth via Euclidean Distance (ED), i.e., MDRHU-ED and Absolute Summation (AS), i.e., MDRHU-AS. The proposed work reduces energy consumption and also provides quality of services.

The literature concludes that it is desirable to predict host utilization in advance to handle dynamic resource requirements and provide quality of service.

3. Support vector regression based host overload detection model

This section contains a detailed description of a machine learning approach: support vector regression for host utilization prediction and algorithm to detect overloaded hosts.

3.1. Problem formulation

We assume a cloud data center with several interconnected hosts having multiple resources. Multiple resource utilization such as CPU, Bandwidth and network utilization are to be included in host utilization evaluation. We are given a input training data set $\{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$ where, $X_i \in R_l$ (Multi-resource usage data), $Y_i \in$ Y (actual utilization value), Y = R, i = 1,2,3…,l, is total number of samples. Using this information, system is to be trained to predict the host utilization with more accuracy.

3.2. Global architecture

A global architecture for the whole process shown in Fig. 1, comprises n number of virtual machines (VMs) and m number of physical machines (PMs) consists of resources: CPU MIPS, RAM, and Bandwidth. The proposed model involves several entities for resource allocation: Cloud data center entity is a collection of heterogeneous computing systems that allows the sharing of resources, provides services to users on-demand, and responsible for communication during allocation. Broker entity works between user and cloud data center and submits the user's service request to cloud data center. Resource Management includes the number of modules, initial allocation, prediction to handle dynamic resource requirements, detection of an overloaded and underloaded host, and on detection, VM migration performed in three steps: VM selection, host selection, and then VM allocation.

3.3. Hybrid kernel support vector regression model (HKSVR)

Support vector machine (SVM) is a statistical learning theory that becomes very popular and evolutionary in machine learning. SVM found to be a prominent approach for classification problems (Cristianini and Shawe-Taylor, 2000). It is easy to classify linear data in a classification problem, but to classify non-linear data, an approach to map non-linear data to higher dimensional data is required. The main attraction towards SVM is of solving non-linear problems by mapping sample to high dimensional feature

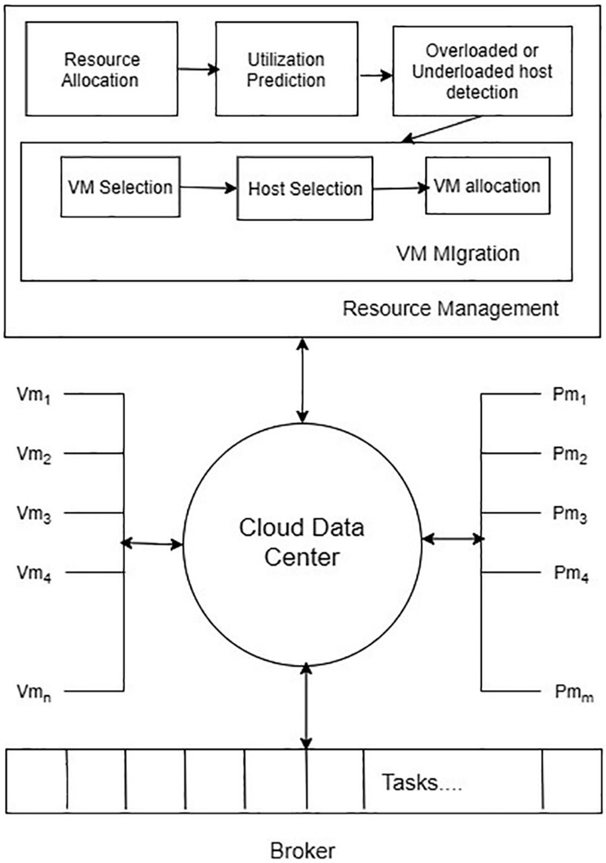


Fig. 1. Global architecture.

space (Vapnik, 1995; Lauer, 2018). Statistical learning theory states that a suitable kernel function can perform the non-linear transform. The Kernel function helps in giving decision boundaries for higher dimensions without complex calculations. In literature, there are four types of kernel functions: Radial Basis Function (RBF), Linear Function (LINEAR), Polynomial Function (POLY), Sigmoid Function (SIGMOID) having their advantages, drawbacks, and shows varied characteristics (Cristianini and Shawe-Taylor, 2000). RBF kernel function, also known as local kernel function, has strong learning ability but weakens generalization performance. Polynomial kernel function, also known as global kernel function, has strong generalization performance. To take advantage of both the kernel function, we have combined RBF and Polynomial kernel functions to improve the support vector machine classifier. According to the kernel function composition condition, the sum of two kernel functions is still a kernel function (Cristianini and Shawe-Taylor, 2000). Therefore, in this paper, we will use hybrid kernel function:

$$K_{hybrid} = \sqrt{\frac{\lambda}{2}} * K_{poly} + \left(1 - \sqrt{\frac{\lambda}{2}}\right) * K_{RBF} \quad (1)$$

where, λ is a constant used to regulate different effects of polynomial kernel function and RBF kernel function, where, $0 \leqslant \lambda \leqslant 1$.

Polynomial kernel functions

$$K(x, x_i) = [(x.x_i) + 1]^d \quad (2)$$

RBF kernel function (RBF)

$$K(x, x_i) = exp\left(\frac{-|x - x_i|^2}{\sigma^2}\right) \quad (3)$$

Let the function map l-dimensional vector x into m-dimensional vector x, where $m > l$ and mapping can be expressed as (Vapnik, 1995):

$$X = ([x]_1, [x]_2, \dots [x]_l)^T \rightarrow ([X]_1, [X]_2, \dots, [X]_m)^T = \phi(x) \quad (4)$$

Let the original input training set is given as: $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ where, $X_i \in R_l$, $Y_i \in$ Y, Y = R, i = 1,2,3…,l is total number of samples.

Next step is to compute decision boundary and separating hyperplane using:

$$f(x) = w.x + b = 0 \quad (5)$$

Maximal margin width of decision boundary reduces error rate and improves accuracy rate. Thus, optimization problem can be written as:

$$max \frac{1}{||w||}$$

$$\text{Subject to}: \quad min|w^T X_n + b| = 1, \quad n = 1, 2, \dots l$$

Can be written as

$$|w^T X_n + b| = y_n(w^T X_n + b) \quad (6)$$

To solve optimization problem a quadratic equation developed as:

$$min\left(\frac{1}{2} w^T w\right) \quad (7)$$

$$\text{Subject to}: \quad y_n(w^T X_n + b) \geqslant 1, \quad n = 1, 2, \dots, l \quad (8)$$

Can be written as:

$$y_n(w^T X_n + b) - 1 \geqslant 0 \quad (9)$$

$$\frac{1}{2} w^T w - \alpha_n(y_n(w^T X_n + b) - 1) \geqslant 0 \quad (10)$$

and Lagrange formulation will be:

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{n=1}^{l} \alpha_n(y_n(w^T X_n + b) - 1) \geqslant 0 \quad (11)$$

On solving Eq. (11) w.r.t w, we get:

$$w - \sum_{n=1}^{l} \alpha_n y_n X_n = 0 \quad (12)$$

$$w = \sum_{n=1}^{l} \alpha_n y_n X_n \quad (13)$$

On solving Eq. (11) w.r.t b, we get:

$$\sum_{n=1}^{l} \alpha_n y_n = 0 \quad (14)$$

To solving w.r.t $\alpha_n \geqslant 0$, substitute Eq. (13) and (14) in (11)

$$L(\alpha) = \sum_{n=1}^{l} \alpha_n - \frac{1}{2} \sum_{n=1}^{l} \sum_{m=1}^{l} \alpha_n \alpha_m y_n y_m (X_n^T X_m) \quad (15)$$

From Eq. (4)

$$(X_n^T X_m) = \phi(x_n^T) \phi(x_m) \quad (16)$$

In form of kernel function Eq. (16) can be written:

$$K(X_n, X_m) = \phi(x_n^T) \phi(x_m) \quad (17)$$

On substituting Eq. (17) to Eq. (15)

$$min\frac{1}{2}\sum_{n=1}^{l}\sum_{m=1}^{l}\alpha_n\alpha_m y_n y_m K(X_n, X_m) \tag{18}$$

Can be written as:

$$min\frac{1}{2}\sum_{n=1}^{l}\sum_{m=1}^{l}\alpha_n\alpha_m y_n y_m K(X_n, X_m) + \sum_{n=1}^{l}\alpha_n \tag{19}$$

Eq. (19) gives decision boundary with maximum margin. With the input training data set, the model trained to calculate decision boundary, classify the data points, and predict the class for future data points.

3.4. Hybrid kernel SVR prediction in cloud environment

To avoid service level agreement violations and improve energy efficiency, it is necessary to handle dynamic resource requirements. Dynamic resource demand requires a system that predicts future utilization based on past utilization history. This research paper proposes a Hybrid Kernel Support Vector Regression (HKSVR) based methodology to predict host utilization using available utilization history. Fig. 2 depicts the steps to train the model and then test it on the given data set to predict the utilization value. Based on the prediction, an overloaded host is detected, and migration is performed to provide Quality of Services (QoS). In this paper, to predict host utilization, first of all, available Alibaba data set (A. Group, 2017) is employed as input. As multiple resources may affect utilization, we have considered multiple resources such as CPU utilization, memory utilization, and Bandwidth utilization for the prediction.

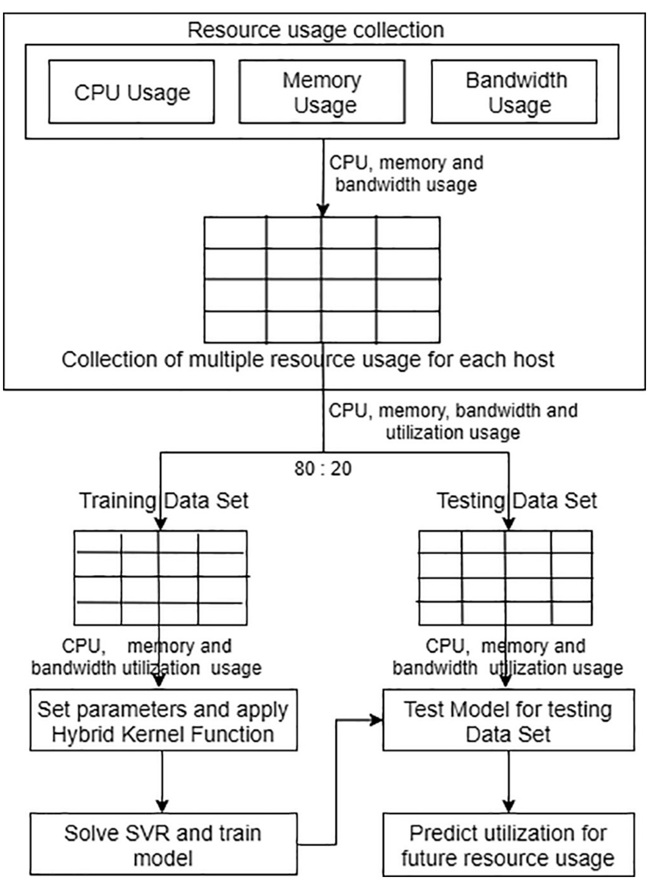


Fig. 2. HKSVR based prediction model.

As shown in Fig. 2, the input data set collects multiple resource usage for each host. The input dataset is split into two parts: training and testing data set. Non-linear functions are mapped to high dimensional data using the proposed hybrid kernel function, set parameters, and then solve SVR equations to train the model. The testing data set is given as input to the trained model and predicts the host's utilization value to test the model. The minimum difference between the actual and predicted value shows the accuracy of the proposed methodology. This HKSVR prediction methodology is implemented in the cloud environment to handle dynamic resource requirements by predicting the host utilization and deciding whether the given host is overloaded or not. On detection of an overloaded host, VM selection process takes place to select the VM for migration followed by VM placement process to place VM on another host.

Algorithm 1. Host Utilization Prediction
Input: CPU utilization, memory utilization, and bandwidth utilization history.
Output: Decision boundary
1 Enter input to the model.
2 Apply hybrid kernel function to map non-linear data to higher-dimensional space.
3 Load training data set (input) and train model to decide boundary line that classifies data.
4 For new data (CPU, Memory, Bandwidth), predict the utilization (P_i)

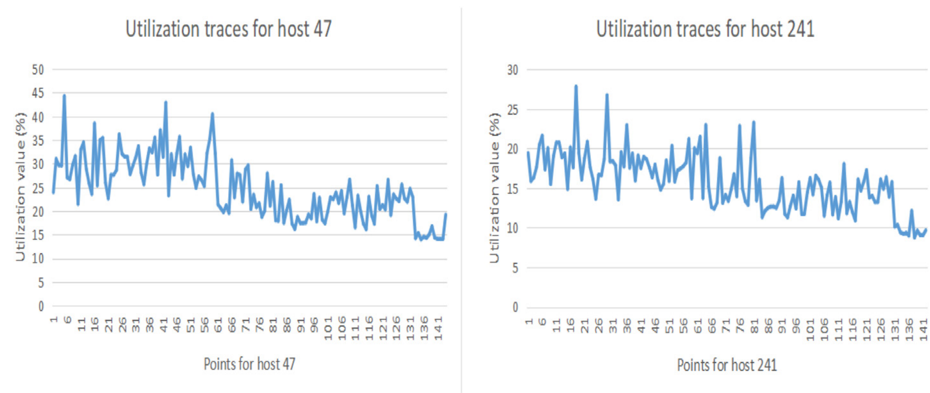
Algorithm (1) describes multi-resource utilization history taken as input and steps taken to train the HKSVR model to decide boundary line.

Algorithm 2. Overload Host detection
Input: HostList, List of their utilization as HostListutilization and decision boundary
Output: Detection of overloaded host
1 for each P_i, $i \in n$
2 decide class for P_i
3 if P_i crosses the decision boundary, the host treated as overloaded
4 VMMigration algorithm performed
5 else
6 does not require any migration for current resources
7 end

Host utilization history is given as input to the algorithm (2). It describes the steps to detect whether the given host is overloaded or not. Similarly, a threshold decided to detect the underloaded host. Once the underloaded host is detected, all the VMs migrated to other hosts to guarantee Quality of Service. In this case, the system will have an overhead of migration cost only.

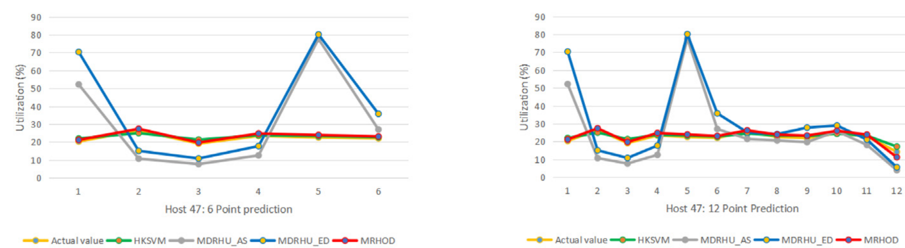
In dynamic consolidation on detecting an overloaded host, VM selection is performed to select a VM for migration. VM migration is performed such that it minimizes energy consumption and migration time. In literature, there exist approaches for VM selection: Random Selection (RS), Minimum Utilization (MU), Maximum Correlation (MC), Minimum Migration Time (MMT), and Constant Fixed Selection (CFS).

1. Random Selection (RS): This approach randomly selects a VM for migration according to the uniformly distributed random variable.



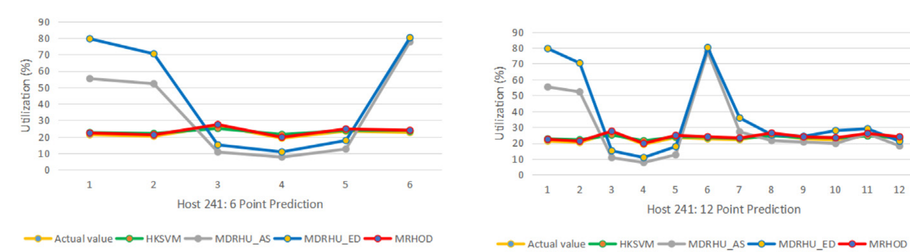
(a) (b)

Fig. 3. Utilization traces on 144 points for host 47 and host 241. (a) Graph for utilization traces of host 47 and (b) Graph of utilization traces for host 241.



(a) 6 points prediction for host 47 (b) 12 points prediction for host 47

Fig. 4. Prediction result of different methods on host 47. (a) Graph for 6 point Prediction on host 47 and (b) Graph for 12 points prediction for host 47.



(a) 6 points prediction for host 241 (b) 12 points prediction for host 241

Fig. 5. Prediction result of different methods on host 241. (a) Graph for 6 points prediction for host 241 and (b) Graph for 12 point Prediction on host 241.

2. Minimum Utilization (MU): VM having minimum CPU utilization selected in this approach to minimize processing overhead.
3. Maximum Correlation (MC): In this approach, probability correlation between resource usage by an application calculated, and VM with the maximum value of this correlation selected for migration.
4. Constant Fixed Selection (CFS): In this approach also, VM randomly selected, but the selection is constant, i.e., the VM either at starting, center, or at last position is selected.
5. Minimum Migration Time (MMT): In this approach (Beloglazov and Buyya, 2012), VM that is going to take minimum migration time selected. Migration time calculated as follows:

$$\text{Migration Time} = \frac{V_m}{B} \tag{20}$$

where V_m is RAM utilized by VM and B stands for bandwidth. So, the VM having this ratio minimum is going to take minimum time in migration.

This paper proposes and implements the approach to predict future resource usage. Based on predicted value, decided whether migration is required or not. On detection of overloaded hosts VMs are migrated to reduce the energy consumption and VM selection done using the above explained MMT approach. After selecting VM, the destination host selected using our proposed earlier paper load-balanced bin-packing approach (Nehra and Nagaraju, under review)

A predefined threshold value is used to detect an underloaded host. On finding any host utilization below that threshold considered underloaded host, VM allocated on that migrated to other active hosts, and switch off the host. Though VM migration is time and energy consuming but the energy consumed in migration is less than the underutilized host's energy consumption.

4. Performance evaluation

This section will discuss some performance parameters to compare prediction approaches, experimental setup, and results and analysis.

4.1. Performance parameters

In this paper, we have compared the proposed HKSVR prediction methodology to other existing prediction approaches. Different metrics (Mason et al., 2018) is employed to analyze the accuracy and performance evaluation. To compare and validate the performance of SVM with other benchmark methods following metrics are considered:

1. Root Mean Square Error (RMSE): RMSE is the standard deviation of the residuals (i.e., the difference of actual and predicted value). It tells how close the data is to the best fit line. RMSE function can be expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - P_i)^2}{n}} \tag{21}$$

where Y_i is the actual value, P_i is the predicted value, and n is the number of observations.
2. Mean Absolute Percentage Error (MAPE): MAPE is a measure of prediction accuracy, used as a loss function in machine learning. Accuracy using this function can be expressed as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - P_i}{Y_i} \tag{22}$$

Y_i is the actual value, P_i is the predicted value, and n is the number of observations, and a low value of MAPE indicates higher accuracy.
3. Mean Square Error (MSE): MSE is a measure of the average of error's square. Here, the error is the difference between estimated and actual value. It is also known as the risk function and reflects the quality measure of the estimator. MSE function expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - P_i)^2 \tag{23}$$

where Y_i is the actual value, P_i is the predicted value, and n is the number of observations. The value returned from the function will always be non-negative and better on closer to zero.
4. R-Squared (R^2): R^2 coefficient of determination is also known as goodness function, it determines how well the regression predicts the actual data points. R^2 function expressed as:

$$R^2 = 1 - \frac{\sum_i (Y_i - P_i)^2}{\sum_i (Y_i - A)^2} \tag{24}$$

where Y_i is the actual value, P_i is the predicted value, A is the average of predicted or observed data. A value closer to 1 indicates perfection to fit the data.
5. Mean Absolute Error (MAE): MAE is the mean of absolute error, the difference between estimated and actual value. The value will always be non-negative. The function to calculate MAE expressed as:

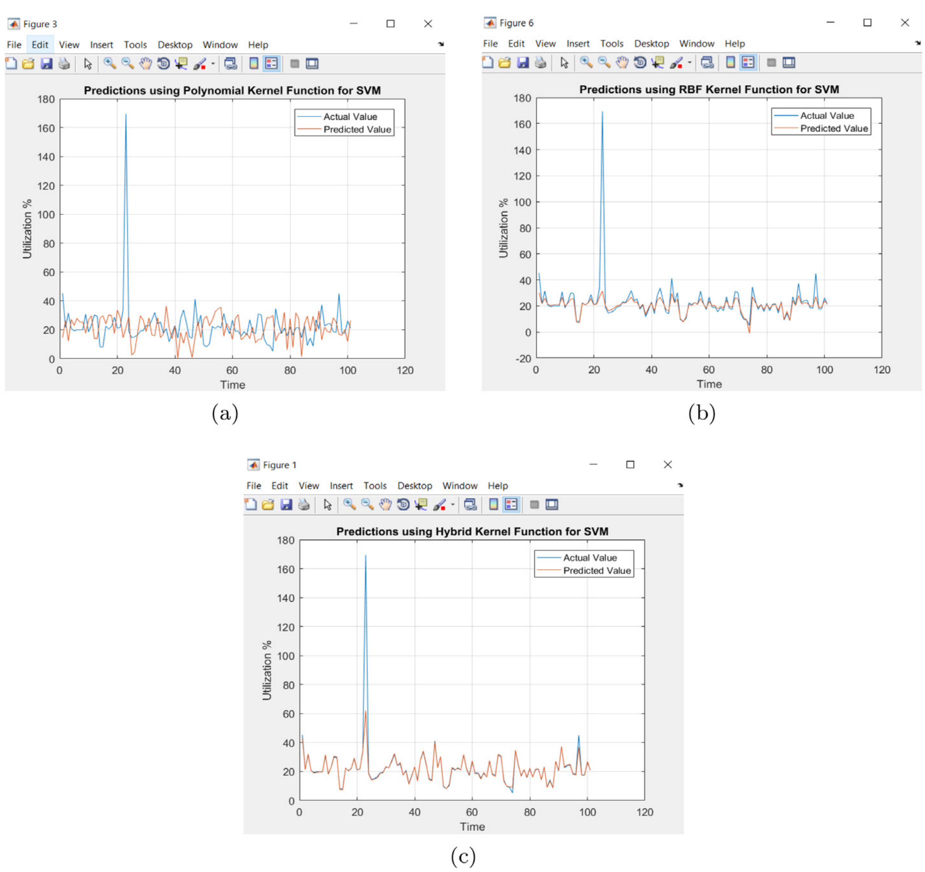


Fig. 6. Comparison of actual and predicted values on using different kernel methods.

Table 2
Comparison analysis of different prediction approaches.

Performance parameter	SVM	MDRHU-AS	MDRHU-ED	MRHOD
RMSE	11.52	12.10	14.66	11.83
MAPE	0.045	1.09	1.02	0.23
MSE	132.81	145	215	133
R^2	0.98	−33	−35.02	1.08
MAE	1.90	6.40	8.78	4.76

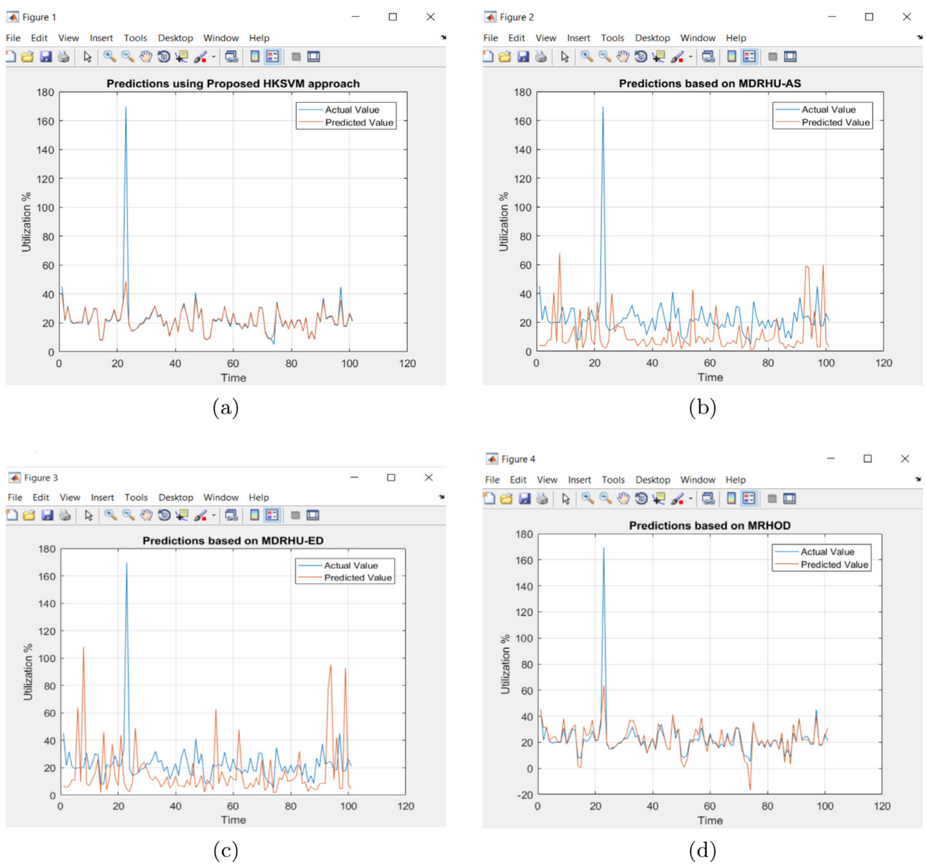


Fig. 7. Graph between actual and predicted value for different prediction approaches. (a) Prediction graph for Proposed Hybrid Kernel SVM (b) Prediction graph for MDRHU-AS (c) Prediction graph for MDRHU-ED and (d) Prediction graph for MRHOD.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(P_i - Y_i)| \tag{25}$$

where, Y_i is the actual value, P_i is predicted value and n is the number of observations.

4.2. Experimental setup and results

In this paper, simulation performed on MATLAB (Levy et al., 2019; MATLAB, 2016) version R2016a. To evaluate performance, proposed Hybrid Kernel SVM based prediction methodology along with existing predictions models (El-Moursy et al., 2012; Ferdaus and Murshed, 2014) are implemented and compared on aforementioned performance measures with system specifications as: CPU Intel Core i5-8265U, 1.80 GHz, RAM 8 GB and the operating system is Windows 10 pro. For experimental purpose, Alibaba dataset (A. Group, 2017) contains the following information: timestamp, machine ID, CPU utilization, Memory Utilization, and Bandwidth Utilization. All utilization values are represented in percent. The dataset includes data for 1300 machines (hosts) for 12 h. There

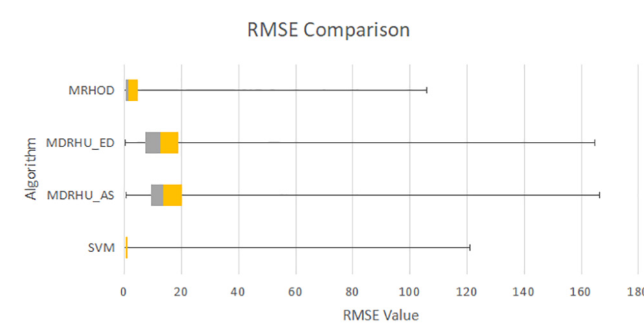


Table 3
Comparison analysis of using different kernel functions for SVM.

Performance parameter	RBF	Polynomial	Hybrid
RMSE	18.44	19.37	11.52
MAPE	0.611	0.72	0.045
R^2	0.95	0.92	0.98
MAE	8.84	10.48	1.90

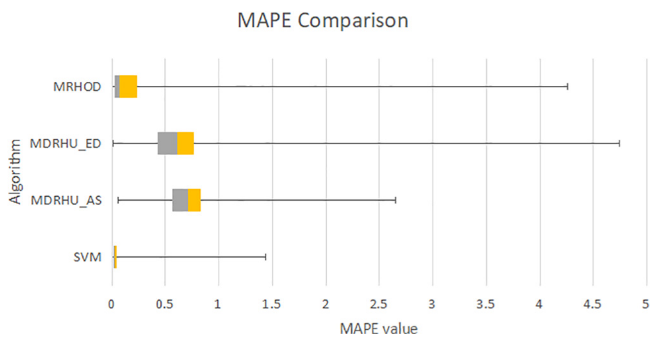


Fig. 9. Box graph for MAPE value of different algorithms.

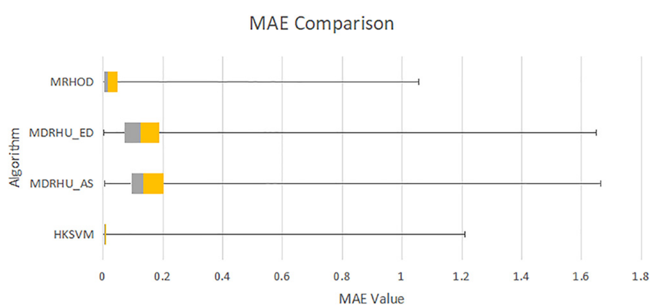


Fig. 10. Box graph for MAE value of different algorithms.

are total 144 number of instances for every machine, each taken on average of 300 s.

The dataset contains 144 points for each host, with 5 min at each point. We randomly selected two hosts: host 47 and host 241, for performance measure on all 144 points. Among 144 points, using 120 points model is trained, and prediction is made for the next 6 and 12 points data.

Fig. 3 shows the utilization traces for host 47 and host 241 for all 144 points. Fig. 4 shows the graph between actual and predicted data for different prediction approach on the next 6 and 12 points of host 47. Fig. 5 shows the graph between actual and predicted data for different prediction approach on the next 6 and 12 points of host 241. In both figures, the prediction using the proposed Hybrid Kernel SVM is nearer to the actual value, indicating more accuracy than other prediction approaches. Support vector regression prediction also depends on the kernel function used to map high-dimensional data. RBF and Polynomial kernel functions are the most widely used function. In this paper, we have used the hybrid kernel function, which is a combination of RBF and Polynomial kernel function as represented in Eq. (1) and used in different applications (Tania and Shill, 2019). Fig. 6 shows a graph of actual and predicted values on training SVR using Polynomial, RBF, and Hybrid Kernel function, respectively. From the graph, we can conclude that kernel function plays a significant role in mapping high-dimensional data and prediction. SVR with the Hybrid kernel function shows better prediction followed by RBF and Polynomial. To evaluate the proposed approach's performance, we have considered the alibaba dataset (A. Group, 2017) consisting of various hosts usage data as input. The data set is divided into two parts: training and testing data set in a ratio of 80:20. 80% value of dataset used to train the model and 20% value of dataset used as testing data set. On evaluating error percent, the prediction error is 16%, 70%, 72%, and 21% for SVM, MDRHU-ED, MDRHU-AS, and MRHOD, respectively, and the proposed approach predicts host utilization with 7%, 64%, and 67% more accuracy than MRHOD, MDRHU-ED, MDRHU-AS approaches, respectively. Table 2 represents the comparison of existing and proposed prediction approaches on given performance parameters. The Table 2 shows that MDRHU-ED performs worst, followed by MDRHU-AS. The low value of MAPE presents higher accuracy, and here, for the proposed approach, the value of MAPE is the lowest. Value of R^2 closer to 1 reflects perfection to fit the data, and from the Table 2, we can see that value for the proposed methodology is closer than existing approaches. The proposed approach outperforms and shows better prediction accuracy, followed by MRHOD, MDRHU-ED, and MDRHU-AS. Fig. 7 shows the graph for four different approaches, plotted on actual and predicted values. As we can see, the difference between the actual and predicted value for Fig. 8(a), i.e., the graph plotted using Hybrid Kernel SVM-based proposed approach is minimum, indicating more accuracy. For the existing approach MDRHU-AS and MDRHU-ED, the prediction is accurate for some points, as shown in Fig. 8(b) and (c). MRHOD gives better prediction accuracy results after the proposed approach than the other two existing approaches, as shown in Fig. 7(d). Here, from graphs, we can conclude that prediction through the proposed SVM methodology is more accurate than the approaches mentioned above. Table 3 represents the comparison of different kernel functions used in SVM. From Table 3, we can conclude that training SVM using the Hybrid Kernel function gives better results in terms of MAPE, RMSE, MAE, and R^2 as compared to RBF and Polynommial kernel function. Fig. 8 shows the box and whisker plot for the RMSE value generated for different prediction approaches. As we can see in the plot, the maximum range value for SVM is higher than MRHOD, but on average RMSE value for Hybrid, Kernel SVM is minimum than MRHOD as its minimum, median, and quartile range less than MRHOD. SVM based proposed approach found better than MDRHU-AS and MDRHU-ED.

Fig. 9 shows the box and whisker plot for MAPE value generated for different prediction approaches. The graph shows that the Hybrid Kernel SVR-based approach outperforms MRHOD, MDRHU-ED, and MDRHU-AS, and its value ranges from nearer zero to 1. A low value of MAPE indicates higher accuracy, and from the graph, SVM reflects higher accuracy because of the lower value. Fig. 10 shows the box and whisker plot for MAE value generated for different prediction approaches. In the graph, we can see that the Hybrid Kernel SVR-based approach outperforms MRHOD, MDRHU-ED, and MDRHU-AS.

5. Conclusion and future work

Resource provisioning and VM consolidation are considered effective methods to handle dynamic resource requirements, enhance resource utilization, and avoid violating service level agreements. Research work proposed an SVR based methodology to predict host utilization in the cloud environment. We have proposed an algorithm to detect overloaded hosts, a VM selection approach, and a VM placement approach. The result section depicts that the SVM based approach outperforms existing: MRHOD, MDRHU-ED, and MDRHU-AS prediction approach in term of RMSE and MAPE and also the difference between actual and predicted value observed for SVM based prediction approach as compared to MRHOD, MDRHU-ED, MDRHU-AS.

There can be many directions for future work; one of the main directions of future work is implementing the proposed approach in a CloudSim simulator and forecasting the host utilization using PlanetLab workload traces. Implement VM selection and place selected VM using Load Balanced Multi-Dimensional Bin-Packing approach. Also, implemented to validate the proposed methodology for new data sets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Hu, F., Qiu, M., Li, J., Grant, T., Tylor, D., McCaleb, S., Butler, L., Hamner, R., 2011. A review on cloud computing: Design challenges in architecture and security.

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., 2010. A view of cloud computing.

Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I. Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Fut. Gen. Comput. Syst. doi:10.1016/j.future.2008.12.001.

Nehra, P., Nagaraju, A., 2019.. Sustainable energy consumption modeling for cloud data centers. In: 2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019.

Jennings, B., Stadler, R. Resource management in clouds: survey and research challenges. J. Network Syst. Manage. doi:10.1007/s10922-014-9307-7.

Ferdaus, M.H., Murshed, M., 2014. Energy-aware virtual machine consolidation in IaaS cloud computing.

Abdelsamea, A., Hemayed, E.E., Eldeeb, H., Elazhary, H. Virtual machine consolidation challenges: a review. Innov. Space Sci. Res. J.

Nehra, P., Nagaraju, A., 2019. Scheduling for resource utilization and load balancing in cloud environment. In: 4th International Conference on Computing for Sustainable Global Development.

Mahrishi, M., Nagaraju, A., 2012. Optimizing cloud service provider scheduling by using rough set model. In: Proceedings of 2012 International Conference on Cloud Computing Technologies, Applications and Management, ICCCTAM 2012.

Khan, M.A., Paplinski, A., Khan, A.M., Murshed, M., Buyya, R., 2017. Dynamic virtual machine consolidation algorithms for energy-efficient cloud resource management: a review. In: Sustain. Cloud Energy Serv.. Principles and Practice.

Amiri, M., Mohammad-Khanli, L., 2017. Survey on prediction models of applications for resources provisioning in cloud.

Dinesh Kumar, K., Umamaheswari, E. Resource provisioning in cloud computing using prediction models: a survey. Int. J. Pure Appl. Math.

Masdari, M., Khoshnevis, A., 2019. A survey and classification of the workload forecasting methods in cloud computing. Cluster Comput., 1–26

A. Group. Alibaba cluster-trace-v2017. url: https://github.com/alibaba/clusterdata.

Gahlawat, M., Sharma, P., 2016. Support vector machine-based model for host overload detection in clouds. In: Proceedings of International Conference on ICT for Sustainable Development, Springer, pp. 369–376.

Abdullah, L., Li, H., Al-Jamali, S., Al-Badwi, A., Ruan, C., 2020. Predicting multi-attribute host resource utilization using support vector regression technique. IEEE Access 8, 66048–66067.

Beloglazov, A., Buyya, R., 2012. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers. In: Concurrency Computation Practice and Experience.

Sharma, O., Saini, H., 2016. VM consolidation for cloud data center using median based threshold approach. Procedia Comput. Sci.

Li, Z., Yu, X., Yu, L., Guo, S., Chang, V. Energy-efficient and quality-aware VM consolidation method. Fut. Gen. Comput. Syst. doi:10.1016/j.future.2019.08.004.

Mahdhi, T., Mezni, H., 2018. A prediction-based VM consolidation approach in IaaS Cloud Data Centers. J. Syst. Software. doi:10.1016/j.jss.2018.09.083.

Li, L., Dong, J., Zuo, D., Wu, J. SLA-aware and energy-efficient VM consolidation in cloud data centers using robust linear regression prediction model. IEEE Access. doi:10.1109/ACCESS.2019.2891567.

Abdelsamea, A., El-Moursy, A.A., Hemayed, E.E., Eldeeb, H. Virtual machine consolidation enhancement using hybrid regression algorithms virtual machine consolidation enhancement. Egypt. Inf. J. doi:10.1016/j.eij.2016.12.002.

El-Moursy, A.A., Abdelsamea, A., Kamran, R., Saad, M. Multi-dimensional regression host utilization algorithm (MDRHU) for host overload detection in cloud computing. J. Cloud Comput. doi:10.1186/s13677-019-0130-2.

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other Kernel-based learning methods.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory.

Lauer, F. Constructing kernel functions. url: https://mlweb.loria.fr/book/en/constructingkernels.htm.

Nehra, P., Nagaraju, A. Efficient resource allocation and management by using load balanced multi-dimensional bin packing heuristic in cloud data centers. In: The International Journal of Universal Computer Science (undetr review).

Mason, K., Duggan, M., Barrett, E., Duggan, J., Howley, E. Predicting host CPU utilization in the cloud using evolutionary neural networks. Fut. Gen. Comput. Syst. doi:10.1016/j.future.2018.03.040.

Levy, M., Raviv, D., Baker, J., 2019. Data center simulations deployed in MATLAB and simulink using a cyber-physical systems lens. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019.

MathWorks, Mastering machine learning.Astep-by-step guide with MATLAB.

Tania, F.A., Shill, P.C., 2019. A modified support vector machine with hybrid kernel function for diagnosis of diseases. In: BECITHCON 2019 – 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health.