International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v9i12.5493
***Article Received**: 17 September 2021 **Revised**: 24 October 2021 **Accepted**: 28 November 2021 **Publication**: 07 December 2021*

Prediction of CPU Utilization in Cloud Environment during Seasonal Trend

Smitha Krishnan
Assistant Professor
College,Changanassery,Kerala
email:nov30smitha@gmail.com

Dr B.G Prasanthi
Assistant Professor SB
StJoseph's College
Bangalore,Karnataka
email:nitai2009@gmail.com

Abstract: Today, the most recent paradigm to emerge is that of Cloud computing, which promises reliable services delivered to the end-user through next-generation data centres which are built on virtualized compute and storage technologies Consumer will be able to access desired service from a "Cloud" anytime anywhere in the world on the bases of demand. Computing services need to be highly reliable, scalable, easy accessible and autonomic to support ever-present access, dynamic discovery and computability, consumers indicate the required service level through Quality of Service (QoS) parameters, according to Service Level Agreements (SLAs) A suitable mdel for the prediction is being developed. Here Genetic Algorithm is chosen in combination with stastical model to do the workload prediction .It is expected to give better result by producing less error rate and more accuracy of prediction compared to the previous algorithm.

Keywords: Cloud Environment, Cost Effective, CPU Utilization etc.

Introduction

Prediction should be cost effective should give less error, should be profitable and should be a success model. It should be accurate, adaptable and based on historic data

The newest proposed approaches are based on machine learning techniques .It predicts the future behaviour, workload, resource provisioning.

Literature survey

Paper	Models	Accuracy Metrices	paramets
[1]	-Holt-Winters	-MAPE,RMSE,PRED(25)	Multi-seasonility
[2]	AR,ARMA,ARIMA,NAIVE	RMSE,MAPE	Autoscaling
[3]	ARIMA	RMSD, NRMSD,MAD MAPE	Qos
[4]	Genetic algorithm	Absolute mean	Seasonal Demand

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v9i12.5493
Article Received: 17 September 2021 Revised: 24 October 2021 Accepted: 28 November 2021 Publication: 07 December 2021

[5]	-	MSE,MAPE	RAM&CPU utilization
[6]	MA ,AR, ARIMA ,DM ,MM	MAPE	Scaling
[7]	MLNN, traditional FLNN	MAE,RMSE	Scaling
[8]	(ARIMA, SVM, FFT, RSLR)	RMSE	Seasonality
[9]	NN,LR	MAPE,PRED(25),RMSE	Auto-scaling
[10]	ARMA,Holt-Winters	MRE,MSE,MAE	Seasonality

Findings :

It is Found That there are not many algorithm to forecast seasonal workload .So an algorithm using genetic algorithm combined with statistical model is used to get better demand function and there by getting greater accuracy and less error while predicting

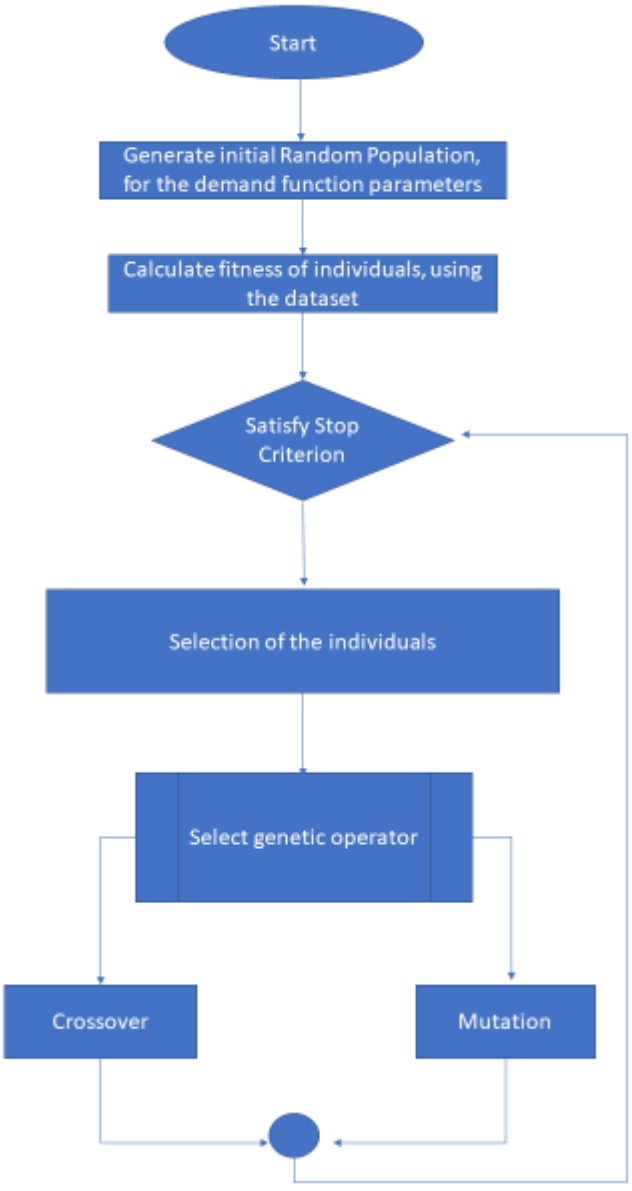
Motivation

Forecasting of workload is important to plan resource allocation. Prediction of all the requirements in advance will help to solve the problem of over provisioning or under provisioning of resources in cloud environment.

Genetic algorithm

There are five phases in a genetic algorithm.

1. Initial population-Every individual is represented either through binary or real values or character based encoding (binary), First population is created randomly.
2. Fitness function-Iterations of population produces generations. Each generation produces a fitness score for individual to complete which represents the ability of an individual to compete. Parents are the fit individuals are allowed to mate to form generation
3. Selection-The fittest individuals are selected fo next generation making. There are many selection methods
4. Crossover-Selected parents are crossed over to produce offspring's using any one of the available cross over techniques.
5. Mutation-Any one of the bit is mutated to bring changes in offspring's



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v9i12.5493
Article Received: 17 September 2021 Revised: 24 October 2021 Accepted: 28 November 2021 Publication: 07 December 2021

Data set used

To do the prediction we are using Rnd data set .The description of the data set is given below.

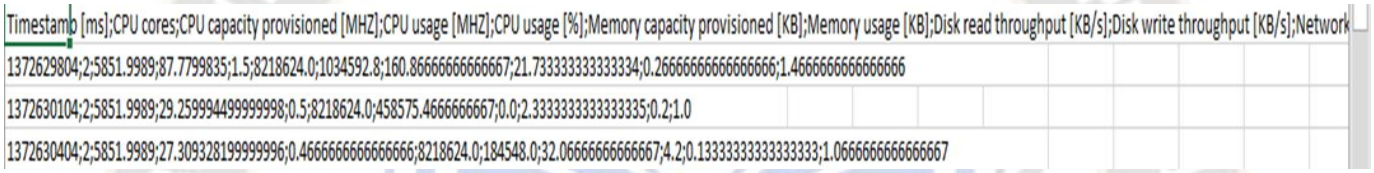
The dataset contains the performance metrics of **1,750 VMs** from a distributed datacentre from **Bitbrains**, which is a service provider that specializes in managed hosting and business computation for enterprises. Customers include many major banks (ING), credit card operators (ICS)etc.

In the Rnd directory, the files are organized into 3 sub-directories by the month that the metrics are recorded.

The format of each file is row-based, each row represent an observation of the performance metrics. Each column of a row is separate by ";\t" The format of each row is

1. Timestamp: number of milliseconds since 1970-01-01.
2. CPU cores: number of virtual CPU cores provisioned.
3. CPU capacity provisioned (CPU requested): the capacity of the CPUs in terms of MHZ, it equals to number of cores x speed per core.
4. CPU usage: in terms of MHZ.
5. CPU usage: in terms of percentage
6. Memory provisioned (memory requested): the capacity of the memory of the VM in terms of KB.
7. Memory usage: the memory that is actively used in terms of KB.
8. Disk read throughput: in terms of KB/s
9. Disk write throughput: in terms of KB/s
10. Network received throughput: in terms of KB/s
11. Network transmitted throughput: in terms of KB/s

Format of the data set included

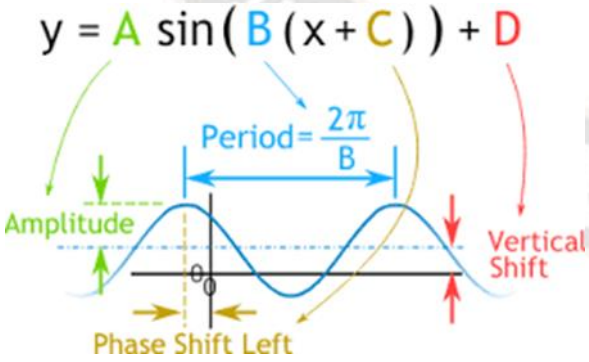


Working of the model

In our model we are going to predict the seasonal demand of our cpu at i^{th} second in $i\text{-}1^{th}$ second and the parameters of the demand function is calculated using genetic algorithm

Demand Function

The demand function is assumed to be the function of time .and it is periodical in nature for seasonal trend. So the periodical wave is sinusoidal in nature and we can represent it with the



We can predict the CPU utilization at 50th sec a second ago itself (i.e. in the 49th sec) by modifying the Demand function

CPU Demand (CD) function = $\frac{A.\sin(B.t_{i-1}+C) + D + Q.t_{i-1}}{U_{i-1}^{e}}$

Where D - Vertical offset

Q-Trend factor
A-Amplitude
B-Freequency
C-Horizontaloffset
CU-Cpu utilization
e-Elascity of cpu utilization

Fitness Function

It is based on relative fitness that a given individual has a probability of entering into the next generation.

*.Fitness function is calculated with the following equation

$$F = \sum_{t} R_t - \frac{(A.\sin(B.t_{i-1}+C) + D + Q.t_{i-1})}{U_{i-1}^{e}}$$

ie $(\text{Demand}_{\text{actual}} - \text{Demand}_{\text{Pred}})$

Where R is the Real cpu utilization at point time t.
F the fitness value

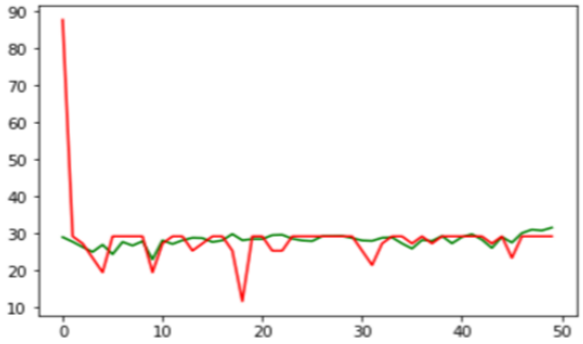
Selection ,mutation and cross over are carried out.

Simulation

crossing_over_probability = 0,1 mutation_probability = 0,3 generation = 500

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v9i12.5493
***Article Received**: 17 September 2021 **Revised**: 24 October 2021 **Accepted**: 28 November 2021 **Publication**: 07 December 2021*

Time



CPU utilization

Adding Regularization to the Fitness Function,
We can add L1 regularization to the fitness function. Regularizing the function help us to get rid of overfitting. As we know our demand function is slightly complex and forms a saw tooth curve, there is a chance that the demand function may get over fitted.
Regularized Error = mean $((Demand_{real} - Demand_{Pred})^2)$ + tuning_param * sum((ind))

Tuning parameter is also known as the learning rate, E.g.: 0.01, which helps in obtaining the global minima.

Conclusion: To build a model that will bring a better accuracy for predicting sesonal demand using regression model combined with genetic algorithm.This model predicts seasonal demand of cpu ahead and so planning in the case of resource utilization becomes easier.As we are using genetic algorithm the calculations are made simpler

References:

[1](IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 11, 2016“Using Multiple Seasonal Holt-Winters Exponential Smoothing to Predict Cloud Resource Provisioning” by Ashraf A. Shahin
[2]”Combining time series prediction models using genetic algorithm to auto-scaling web applications hosted in the cloud infrastructure “ by -valterrog´eriomessias · juliocezarestrel
[3]Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications’ QoS by Rodrigo N. Calheiros, Enayat Masoumi, Rajiv Ranjan, Rajkumar Buyya
[4]”Geneic Algorithms in Seasonal Demand Forecasting” by Chodak, Grzegorz and Kwaśnicki,
[5]2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) “Cloud Workload PredictionUsing ConvNet And Stacked LSTM” by Peyman Yazdanian Saeed Sharifian,Tehran-Iran
[6]2016 IEEE International Conference on Cloud Computing and Big Data Analysis “Workload Prediction for Cloud Computing Elasticity Mechanism “ by Yazhou Hu, Bo Deng, Fuyang Peng Dongxia Wang
[7]2018 IEEE 11th International Conference on Service-Oriented Computing and Applications-”A Resource Usage Prediction System Using Functional-link and Genetic Algorithm Neural Network for Multivariate Cloud Metrics “ by Thieu Nguyen, Nhuan Tran, Binh Minh Nguyen ∗ Giang Nguyen
[8]”CloudInsight: Utilizing a Council of Experts to Predict Future Cloud Application Workloads” by In Kee Kim Wei Wang Yanjun Qi Marty Humphrey
[9]Empirical prediction models for adaptive resource provisioning in the cloud Sadeka Islam , Jacky Keung Kevin Lee , Anna Liu
[10]2017 IEEE 10th International Conference on Service-Oriented Computing and Applications”A Type-Aware Workload Prediction Strategy for Non-Stationary Cloud Service” by JunGuo
11]Workload prediction in cloud environment by smitha krishnan,B.G prasanthi