Survey on Cloud Load Prediction using Machine Learning Techniques

Mrs. Roopa R. Gaur[1], Mrs. Sushama C. Budruk[2], Mrs.Nutan V. Patil[3]
[1,2,3]*Member, NanasahebMahadik College of Engineering*

***Abstract*—Cloud computing is technology which allows on demand accessibility of different computer resources, computing power and data storage without active presence of user. For any application Resources provisioning is main important challenging work in cloud computing environment.Resource need to be allocated dynamically in cloud platform, on the basis of workload behavior of different applications. Any alteration, in resources provisioning it may leads waste of energy, storage and cost. In Addition to, it creates Service Level Agreements (SLA) transgression and dropping of Quality of Service (QoS).Hence any violation to commitment made for SLA the service provider has to pay penalty and it leads to customer dissatisfaction. Hence workload prediction of cloud environment is necessary for this purpose there are different technology to predict cloud work load, in this paper we have proposed different machine leaning techniques such as Linear regression, SVM and artificial neural network which creates magic in predictions of workload. Cloud computing along with machine learning could leads to more benefits.This paper presented characteristics and importance of several prediction schemes to enhance workload management system in cloud environment.**

***Index Terms*—Cloud Computing, Workload, Prediction Model, Quality of Service, Service Level Agreement.**

I. INTRODUCTION

Machine learning is one of the application artificial intelligence (AI).It makes system to learn automatically with study of algorithms which have ability to learn through different patterns based on that it can give better perditions. The application of cloud involves storage, computing and networking. Cloud with features of machine learning increases the capability of cloud. It learns from huge amount data stored in cloud, and it provides the predictions and analysis.

The cloud computing has made modern day enterprise proprietors (with restrained capital for instance) to hire and use the infrastructure or resources needed to run their teams in a cost-effective manner as you use. Within the concept of a Cloud Computing provider, the workload pattern, cloud-based feeds (public, private, mixed, and network), and cloud provider provider models (SaaS, PaaS, and IaaS) are carefully integrated individually. From a cloud-based corporate perspective, Cloud loads incoming work in line with static, periodic, unpredictable, flexible loads, and fast lifetime loads. Stable and periodic workloads are usually based on a predictable sample size of their arrival. Ongoing work shifts reflect a sample of diversity that is reflected in the general growth and decline of fashion in the frequencies of their arrival. Unpredictable workloads reflect random array of frequent arrivals and are the most difficult type of predictive test tasks. As soon as a lifetime job is a rare burden to arrive and its shipping is greatly appreciated by customer support. [2] A cloud SLA (a cloud service level agreement) is an agreement between a cloud service company and a client that ensures a minimum level of service is maintained. Ensures availability, reliability, and responsiveness levels in systems and applications. An SLA is a warranty between the provider and the client.

Predicting the availability of new functionality in computer network systems is a fundamental problem from a cloud computing. Making such predictions is not easy because of the flexible nature of current computer programs and their function. Ensuring high efficiency, flexibility, and cost-effectiveness and environment for cloud platforms needs to be able to quickly plan and deliver services, which will ensure that the supporting infrastructure can be closely aligned to the needs of the various applications.

Machine learning is such amazing area where using different algorithms of machine learning one can easily predict the load on cloud such as CPU

utilization, memory usage etc.[8]This paper presents different prediction algorithms such as Linear regression ,SVM, Bayesian models and artificial neural networks, which are used for cloud applications. Rest of the paper is framed as follows: Section2 characteristics Section 3 presents various machine learning techniques. Section 4 describes comparisons. Section 5 presents final conclusion.

2. CLOUD COMPUTING

Cloud computing is the delivery of computer offerings comprising servers, storage, websites, network, software, statistics, and online intelligence ("cloud") to provide faster innovation, curved assets, and scale economy.

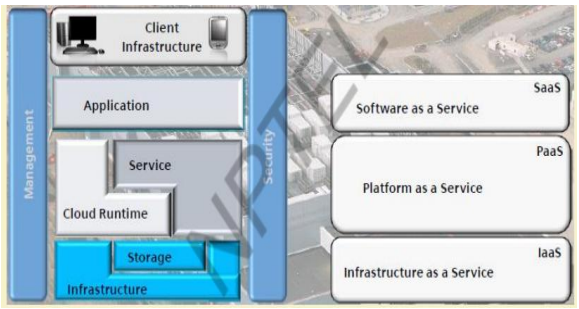


Fig 1: Cloud Computing Architecture

Essential Characteristics of Cloud are as follow:

1. Integration of Services: This enables the cloud service provider to provide its customers with a multi-employer model. Visual and practical services are allocated and redistributed according to customer needs.
2. Custom Customization Service: Means that the client (client) can use different cloud services as required without the involvement of the cloud.
3. Limited service provider: Cloud services are managed and monitored with the help of a cloud provider. This is important in payment; obtain management permits, resource utilization, power generation programs and other obligations.
4. Network Access: Cloud provider capabilities are available on the network
5. Quick Expansion: Described as the ability to measure resources both up and down as needed.

2.1. Cloud Workloads

Full cloud functionality reaches Cloud datacenters in the form of user-driven tasks.All work includes some self-explanatory features such as delivery time, user ID, and its corresponding service requirements according to CPUandmemory.Asinglejob may contain one or more functions, i.e..is scheduled to be processed on Cloud servers. One task may require one or more processes.

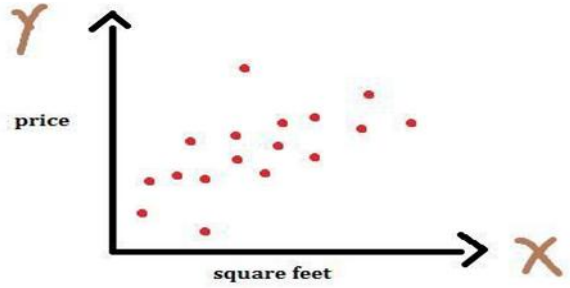
3. MACHINE LEARNING TECHNIQUES

Machine learning provides computers with the ability to learn without being explicitly programmed. It takes the data and learns about it. Using different algorithms it analyses data and forecast the next occurrence or future possible value of data. The different machine learning algorithms are given below.

3.1Linear Regression

It is a machine learning algorithm based on supervised reading where the name of the label is already given. Perform a rewind function. Going back to the target prediction / output value model based on the given independent variables. It is widely used to find relationships between fluctuations and predictions / predictions. Different reversal models vary based on - the type of relationship between dependent and independent variables, the processing and the number of independent variables used [4].

A line drop is a line model, e.g. A model is provided that captures the linear relationship between input variables (x) and single output variations (y). Specifically, that y can be calculated from the line combination of input variables (x). In a simple line rotation, we create a relationship between the output / target variation and the given input variation by inserting a line, known as a regression line.



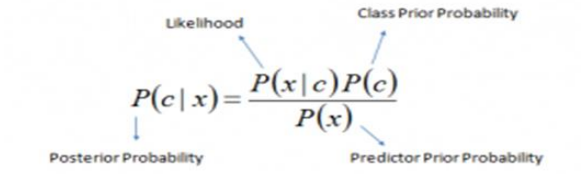
Generally, a line can be given with the line number y = m * X + b + ∈. There, y is a dependent variable, X is an independent variable, m is a slope, b short. In machine learning, we can rewrite our equation with y (x) = w0 + w1 * x where w is the parameters of the model, x input, and y is the direct variable. Where,

Y-Dependent

X-Independent
b-Intercept
m-Slope
∈ - Error
If a single input variable (x), the method is called to as simple linear regression.When there are multiple input variables, the method is multiple linear regressions.

3.2 Bayesian model
Naive Bay's Classifiers is primarily based on the Bayes' Theorem, that's based on conditional probability or in simple terms, the probability that an event (A) will take place given that any other event (B) has already occurred. Essentially, the theory allows a hypothesis to be updated whenever new proof is brought[5]. The equation under expresses Bayes' Theorem inside the language of probability:

Problem statement:
- Features given X1, X2,…,Xn
- Guess the Y label
X = (Rainy, Hot, High, False) and y = No



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

• P (c | x) posterior class probability (c, targeted) predicted (x, attributes).
• P (c) pre-class probability.
• P (x | c) chances are probability of a given class of prediction.
• P (x) predictor probability.

3.3 Support Vector Machine
Support Vector machine is yet powerful Supervised machine learning algorithm that may be used for constructing both regression and classification models. SVM algorithm can perform genuinely properly with each linearly separable and non-linearly separable datasets. in spite of a constrained quantity of statistics, the support vector machine algorithm does no longer fail to reveal its magic.Support vector machine algorithm is based on the concept of 'decision planes', where hyper planes are used to classify a set of given objects[6].

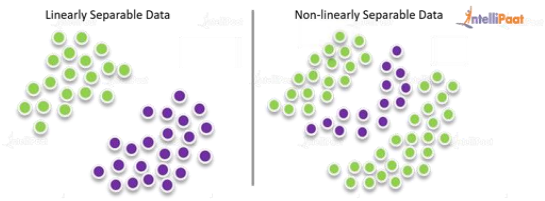


Fig 2: Linearly Separable and Non-linearly Separable Datasets (SVM)

3.4 Artificial Intelligence (ANN)
Neural Network is used for forecasting, which predicts future load based on historical data [7]. It is a machine designed to model how the brain performs a specific task ANN represents the Neural Networks performed. Basically, a computer model. That is based on the structures and strengths of biological neural networks. Despite the fact that, the formation of ANN is influenced by the flow of information, the changes in the neural network were based entirely on inputs and outputs. It contains a large number of interconnected processing elements called neurons to perform all functions. Information stored in neurons is a heavy link of neurons Input signals reach objects processing through connections and connecting weights. It has the ability to read, remember and integrate given calculations through proper allocation and weight correction

3.4.1 How simple neuron works?
Suppose there are two neurons X and Y that can transmit signal to all other Z neurons. Then, the X and Y are input neurons and the Z is an outgoing neuron. Input neurons are connected to an outgoing neuron, over the connective tissue (A and B) as evidenced by the discern.
Here the example of neural network sample load of data center is analyzed and given as input

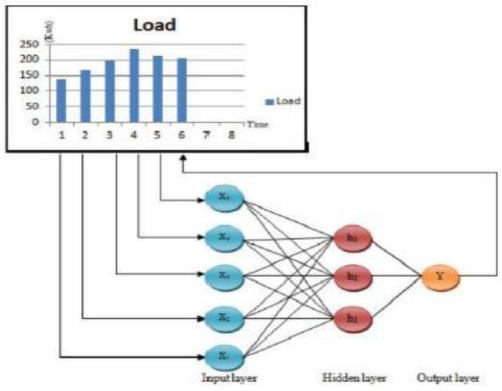


Fig 3: neural network analysis

IV. LITERATURE SURVEY

In this paper the author proposes how to predict server full performance based on machine learning using data from a database based on authentic database, large scale, and business class. They also calculated the failure rate of the failure and the warning level of the models for overcrowding, and as suggested by the test supports the cost of processing the overcrowding. Test results show that machine learning methods especially Random Forest can better predict server load than the standard mathematical analysis method. Compared with the conventional method of statistical analysis, the proposed method uses smaller data and lower magnitude, and produces more accurate predictions. [1]

A service level agreement (SLA) is an important part of cloud systems to ensure maximum availability of services to clients. For a breach of the SLA, the provider must pay fines. In this paper, the authors examined two machine learning models: Naive Bayes and Random Forest Classifiers to predict SLA violations. As SLA violations are a rare occurrence in the world (~ 0.2%), the task of segregation becomes difficult. To overcome these challenges, they used a few methods to repeat the samples. They found that random forests with a new SMOTE-ENN sample have easy operation among other methods with accuracy of 99.88 to F_1 points of 0.9980. [8]

In this paper, a different method is proposed for the prediction of the future load for cloud-based data centers. This method works in different phases[9]. Initially, Bayesian model is utilized to anticipate the mean burden over a long term time interval which is contrasted and PSR and EA-GMDH technique which joins the Phase Space Reconstruction (PSR) strategy and the Group Method of Data Handling (GMDH) for successful expectation then Neural Network predicts the future load dependent on the previous recorded information which separates itself with the nearness of covered layers followed by support vector and kalmann smoother which is a multi-stride ahead CPU load forecast technique dependent on Support Vector Regression which is entirely steady, for example its expectation mistake increments gradually as the anticipated advances increment.

This paper proposed a self-versatile expectation suite with an intend to improve the exactness of predictive auto-scaling frameworks for the IaaS layer of distributed computing. The expectation suite utilizes the choice fusion technique and facilitates the choice of the foremost accurate prediction algorithm and therefore the window size with reference to the incoming workload pattern[10]. The proposed architecture used the strategy and therefore the design patterns which guarantee the automated runtime selection of the acceptable prediction algorithm also as detection of an appropriate workload pattern and an appropriate window size.

In this paper prediction method supported Bayes model is proposed to predict the mean load over a long-term interval, also because the mean load in consecutive future time intervals. This method [11] identifies novel predictive features of host load that capture the expectation, predictability, trends and patterns of host load. It also determines the foremost effective combinations of those features for prediction. Tests show that the Bayes method reaches high correctness with a mean squared error of 0.0014. Also, the Bayes method develops the load prediction correctness by 5.6-50% compared to other state-of-the-art methods supported moving averages, auto-regression, and/or noise filters.

In this paper-based forecast method based on the Bayes model it is proposed to predict the long-term load value, and also because the average volume for future consecutive periods. This approach [11] identifies new predictive host load features that capture expectations, predictions, styles and load load patterns. It also determines the most effective combinations of those predictive factors. Tests show the Bayes approach achieves maximum accuracy with a 0.0014 square measure error. Also, the Bayes method improves the accuracy of load prediction by 5.6-50% compared to other modern methods supported by motion measurement, automatic rotation, and / or sound filters.

V.CONCLUSION

This paper presents a survey on cloud workload prediction with machine learning techniques. In the beginning it presents importance of SLA what the difficulties are raising while providing resources such as CPU, memory usage, servers etc. to particular application. SLA sets expectation for both cloud service provider and client. In order to achieve

effective SLA, cloud service provider has to know proper utilization of resource even in case any new request by client. By predicting the expected workload to service provider and allocating appropriate levels of resources to execute the user requests one can use machine learning techniques for excellent and efficient prediction

REFERENCES

[1] S.Chen, B.Mulgrew and P.M.Grant, ―Aclustering technique for digital communications channel equalization using radial basis function networks,‖ *IEEETrans. On Neural Networks*, vol. 4, pp.570-578, July 1993.

[2] J. U. Duncombe, ―Infrared navigation—Part I: An assessment of feasibility, ‖*IEEE Trans. ElectronDevices*,vol.ED-11,pp. 34-39,Jan.1959.

[3] C.Y.Lin,M.Wu,J.A.Bloom,I.J.Cox,andM.Miller, ―Rotation, scale, and translation resilient publicwatermarkingforimages,‖*IEEETrans.ImageProcess.*, vol. 10,no.5,pp.767-782, May2001.

[4] A.CichockiandR. Unbehaven, *Neural Networks forOptimizationandSignalProcessing*,1sted.Chichester,U.K.:Wiley,1993,ch.2,pp.45-47.

[5] W.-K.Chen, *Linear Networks and Systems*, Belmont,CA:Wadsworth,1993, pp. 123-135.

[6] H.Poor,*AnIntroductiontoSignalDetectionandEstimation*;NewYork: Springer-Verlag, 1985,ch.4.

[7] R.A.Scholtz,―TheSpreadSpectrumConcept,‖in *MultipleAccess*, N.Abramson, Ed. Piscataway, NJ:IEEEPress, 1993,ch.3, pp. 121-123.

[8] G.O.Young,―Syntheticstructureofindustrialplastics,‖ in *Plastics*, 2nd ed.vol.3, J.Peters, Ed. NewYork:McGraw-Hill,1964,pp.15-64.

[9] S. P.Bingulac,―On the compatibility of adaptive controllers,‖in*Proc.4thAnnu.AllertonConf.CircuitsandSystemsTheory*,NewYork,1994, pp.8-16.

[10] W. D. Doyle, ―Magnetization reversal in films with biaxial anisotropy, ‖in*Proc.1987 INTERMAG Conf.*, 1987,pp. 2.2-1-2.2-6.

[11] G. W. JuetteandL. E. Zeffanella, ―Radio noise currentsnshort sections on bundle conductors, ‖presentedattheIEEESummerPowerMeeting,Dallas,TX,June 22-27,1990.

[12] J.Williams, ―Narrow- bandanalyzer, ‖Ph.D. dissertation, Dept. Elect.Eng., Harvard Univ., Cambridge, MA,1993.

[13] N. Kawasaki, ―Parametric study of thermal and chemical nonequilibriumnozzle flow,‖ M.S. thesis,Dept.Electron.Eng.,OsakaUniv.,Osaka,Japan, 1993.

[14] J.P.Wilkinson, ―Nonlinear resonant circuit devices, ‖ U.S.Patent3624 12,July16, 1990.

[15] *LetterSymbolsforQuantities*,ANSIStandardY10.5-1968.

[16] *TransmissionSystemsforCommunications*,3rded., WesternElectricCo.,Winston-Salem,NC,1985, pp. 44-60.

[17] *Motorola Semiconductor Data Manual*, Motorola SemiconductorProductsInc.,Phoenix,AZ, 1989.

[18] R.J.Vidmar.(August1992). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. 21(3). pp. 876-880. Available: http://www. halcyon. com/pub /journals/21ps03-vidmar