



A WEIGHTED ENSEMBLE OF AUTOMATIC ALGORITHMS FOR VIRTUAL MACHINE PERFORMANCE PREDICTION IN CLOUD

Sudha V¹, Archana M², Chaithra D H³, G M Chethana⁴, Kavya S⁵

¹Asst. Prof at Dr.TTIT, ^{2,3,4,5}Student at Dr.TTIT

¹Sudhaviju@gmail.com, ²ammuammu1129@gmail.com, ³chaithrachaitu097@gmail.com,

⁴g.m.chethana97@gmail.com, ⁵kavyaachu.s@gmail.com

Abstract

In cloud computing virtual machines are hosted on a shared computing platform, which compete for resources such as CPU, I/O and Network Storage. Predicting resources usage in cloud environment is a challenging problem due to varying characteristics of workloads running on cloud. Each virtual machine resource usage can be characterized as steady, trend, seasonal, cyclic or bursty pattern. Manually it is not feasible to fit the Predicting models for each of thousands of VMs running on cloud. Further different Predicting models are suitable for different type of workload. Manual selection of best model is will take too much time and complicated. In this manuscript we propose, implement and evaluate an automated technique of combining the Predictions from multiple non-overlapping time series Predicting methods. The objective is to improve accuracy and robustness of the predicting algorithm. The results show that proposed technique is successful in Predicting with better accuracy, irrespective of workload type and without any manual intervention. **Keywords:** Cloud computing, Ensemble Forecasting Time series forecasting.

[1] I. INTRODUCTION

Cloud computing is amalgamation of service oriented computing and utility computing paradigms where both hardware and software are provided as-a-service model [1]. In cloud a subscriber can rent a service, increase or

decrease resources allocated and are billed according to pay-per-use model. This dynamic, shared and elasticity characteristics of cloud makes cloud service performance unpredictable and vary over time which may result in poor subscriber experience and bad reputation of the cloud service provider(CSP). The CSPs are facing the biggest challenge of meeting the negotiated Service Level Agreements (SLAs) to the subscriber.

A self-adaptive prediction method using ensemble model and subtractive-fuzzy clustering based fuzzy neural network (ESFCFNN) is proposed for performance prediction in cloud [5]. In this work an ensemble model is created using base predictors such as Auto Regressive (AR), Trend Seasonality Model (TSM), and Moving Average (MA). The combined results are sent to fuzzy neural network with self-adjusting learning rate and momentum weight. To further optimize convergence, fuzzy subtractive clustering algorithm is proposed. The effort in select few models that work well with uni-variate and multivariate time series concurrently. In common random variables are uncorrelated and extremely dependent. And in case of random vector, a multivariate normal distribution, two or more of its constituents which are couple wise independent are independent and it is to be noted that, it is not true that two random variables which are generally distributed and uncorrelated are independent. Strictly, dependence denotes to any condition with which random variables cannot fulfil a mathematical condition of independence. Correlation could be reflected as any parting of

two or more random variables from independence, and technically it denotes to any of several additional specialized types of relationship between mean values. There are many such instances for example in 2014 many CSPs like Drop box, Google, Adobe etc experienced service outages lasting hours to weeks[2].

The primary causes for outages are unanticipated hardware, software or network faults. The secondary causes which lead to first cause are unpredictable performance due to workload variation, a software bug, security attacks and more. 'Resource pool manager' which manages the resources in term that refers to the grouping together of resources (CPU usage, memory usage, network bandwidth usage etc.) for the purpose of maximizing the advantage and minimizing the risk to the users. This Resource pool manager utilizes a well-organized load balancing algorithm to poise stock on every cloud service provider in addition to optimal cloud utilization cost. By means of this prediction model, the conclusion to assign and discharge virtual machine could be completed positively, as a result, creating cloud effective by both economy and performance. Thus we propose an algorithm that combines the forecasts from set of base automatic forecasting models. This has advantage that it does not require domain knowledge of expert to use the algorithm. Further combining forecasts improve accuracy, adaptability, robustness against changes in the workload pattern in cloud environment. A collection of methods that are alike in nature will perform more inadequately than a more assorted model set. Ensemble-learning is a machine learning model where numerous users are trained to resolve the similar problems and it makes better than single learning method and ascertains regularities in vital. In this manuscript the methodology utilizes weighted ensemble approach, which allocates the weights for various models depending on the results i.e. Predicted values & RMSE value.

Such time-series data can be used to train statistical forecasting models and used predict future data. The workloads running on cloud have diverse characteristics viz. steady, increasing or decreasing trend, seasonal, cyclic, bursty or random [4]. There are many important

time series forecasting algorithm such as Auto Regressive Integrated Moving Average (ARIMA), Exponential smoothing state space models (ETS), Neural Network etc which can be used to forecast the future resource usage values.

In many circumstances, it is advantageous or essential to alter a time series data set before using the refined methods. As mentioned in this paper, it has non-stationary data, a suitable preliminary transformation of data to dig up stationary may do well in stabilizing the discrepancy and it may use one of the common time series models. Although if some models could be functional straight to non-stationary time series without need of a preliminary transformation of the data, However the case study reveals, uni-variate functions could be applied point way to multivariate data to alter their marginal distributions. It is achievable to modify some attributes of a multivariate distribution by the use of appropriately constructed transformation.

II. LITERATURE REVIEW

In this section we discuss the previous work on ensemble forecasting techniques used in cloud, grid and virtualized environment.

A self-adaptive prediction method using ensemble model and subtractive-fuzzy clustering based fuzzy neural network (ESFCFNN) is proposed for performance prediction in cloud [5]. In this work an ensemble model is created using base predictors such as Auto Regressive (AR), Trend Seasonality Model (TSM), Moving Average (MA). The combined results are sent to fuzzy neural network with self-adjusting learning rate and momentum weight. To further optimize convergence, fuzzy subtractive clustering algorithm is proposed. The authors conclude that the system is complex and aim to improve its efficiency [5].

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \theta_q e_{t-q} + e_t$$

A Multi-step-ahead load forecasting technique (KSWSVR) integrates support vector regression and kalaman smoothing filter. The kalaman smoothing filter is used for preprocessing and SVR for prediction. This approach uses regression and does not consider the varying

nature of the workload i.e Seasonal, Trend, cyclic and Random [6].

A selfadaptive prediction system for forecasting virtual machine (VM) demand in cloud is proposed in[7]. It implements an ensemble of MA, AR, ANN, SVM and gene expression programming.

Compared to our approach we use apt number of base predictors hence expected to be more efficient and accurate. In [8] an ensemble of Holt-Winters, ARIMA and StructTS base predictors are used.

The forecasts are combined using a weighted k-nearest-neighbor algorithm. Here it does not address the varying nature of the workload.

In [9] uses fast fourier transform (FFT) and Markov chain for resource prediction. The FFT is used to predict periodic cycle service requirement and markov chain is used to predict a periodic service request. The accuracy is compared with only ARMA approach.

The approach used in [10] forecasts cloud workload using AR, Holt –Winter, Markov Chain, average ensemble, ensemble using 3 layer FFNN. The results show that average ensemble model which takes average of forecasts produced by each model performs better compare to individual forecasts and combined forecasts using 3 layer FFNN. In [11] ANN, SVM and ensemble of ANN models are used for modeling the performance of VM hosted applications. The results show that ensemble technique improves the performance. The previous work shows that ensemble forecasts provide promising direction for improving accuracy of forecasts in cloud.

Further our approach is different from previous work in following directions:

1) Uses minimal set diverse non-overlapping base predictors. By reducing the number of base predictors we expect to improve efficiency.

2) Filters the high error predictors to improve accuracy.

III. PROPOSED SYSTEM DESIGN

The proposed ensemble system design as shown in fig 1. It uses diverse combination of base predictors capable of handling seasonal data, non-seasonal data, linear and nonlinear data. To keep the model simple and lightweight we restrict the number of base predictors to four i.e

Automatic ETS, Automatic Arima, Neural Network and Trigonometric Box-Cox ARMA Trend Seasonal Model (TBATS).

- Automatic ETS: capable of forecasting trend and/or seasonal data [12].
- Automatic Arima : capable of forecasting cyclic patterns and short term trends [12].
- Neural Network regression: capable of forecasting complex seasonal and trend components [13].
- Tbats: capable of forecasting complex seasonality [14].

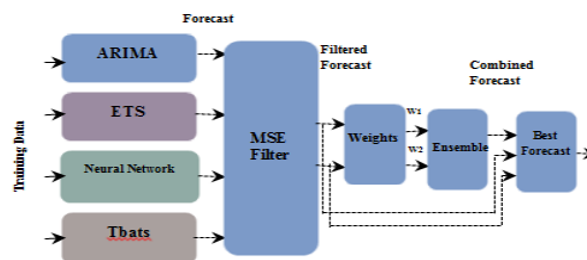


Fig. 1. Proposed System Design

All of the chosen models are further explained briefly here.

A. Automatic ETS: Exponential smoothing (ETS)[15] predictions are weighted average of past observation with weights decaying exponentially as observations get older. ETS Predicting model takes error, trend and seasonal components into consideration.

Hyndman et al used the automatic ETS model to the M-Competition data and IJF-M3 competition data and suggested that it is good for short term prediction and seasonal short-term time series. ETS models are suitable for handling non-linear data but assume the data is non-stationary. A non-stationary time series is one whose properties depend on time at which the series is observed. The trend and seasonality affects the value of time series at time of observation therefore are nonstationary. The exponential smoothing models are considered best to handle seasonal data.

B. Automatic ARIMA: ARIMA focuses on autocorrelation in data i.e., correlation between elements of same series separated by a given interval. Whereas ETS considers error, trend and seasonality features of time

series, ARIMA is basically for stationary time series[15]. A stationary time series is one which depends on the time at which it is observed and its mean and variance remains fixed over time. For non-stationary time series it requires differencing the time series first. Thus, ARIMA is combination of Autoregressive (AR) and Moving Average (MA) models along with differencing/Integration on non-stationary time series. The Auto regression (AR) is Predicting variable of interest using linear combination of past values of variable. The moving average (MA) model is using past predict errors in regression like model rather than past values of variable. An ARIMA model is denoted as ARIMA (p,d,q). Where p is order of auto regressive part, d is degree of first differencing and q is order of moving average part.

Where y = the predicting variable, c = constant, ϕ = coefficient of parameter p, θ = coefficient of parameter q, e = errors in time t.

C. Nonlinear Neural Network (NN): The Nonlinear Neural Network model used in the proposed work is system with feed-forward neural network with lagged inputs and one hidden layer[16]. The nnetar function of R trains 25 neural networks starting with random initial values and then takes average of predictions to compute predictions [13]. The feed forward neural network with one hidden layer takes following form.

$$\hat{y}_t = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j \varphi(x'_t \cdot \hat{y}_j)$$

Where x_t consists of p lags of y_t and the function Ψ is of logistic form.

$$\varphi(x'_t \cdot \hat{y}_j) = [1 + \exp(-\hat{y}_{j0} + \sum_{i=1}^p \hat{y}_{ij} \cdot y_{t-1})]^{-1} s$$

Where $j=1, \dots, k$
The nonlinearity will happen through lagged y_t in the logistic function. The k value is number of logistic function known as number of the hidden nodes. The parameters of

Neural network are selected based on loss function embedded in learning algorithm. The loss function could be for example the Root Mean Square Error (RMSE).

D. Trigonometric Box-Cox ARMA Trend Seasonal Model (TBATS): The TBATS model is used for Predicting with complex seasonality data [17]. TBATS is model of type exponential smoothing state space model with Box-Cox transformation, ARMA error correction, global and local Trend and Fourier-like seasonal components

IV. ENSEMBLE LEARNING APPROACH
The proposed approach first splits the data into training set and validation set. Then it trains each model on the training set and generates the forecasts. Then it trains each model on the training set and generates the forecasts

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \dots\dots\dots$$

$$\dots\dots\dots(1)$$

The models which do not fit the data well tend to produce high error i.e high MSE. Hence we will eliminate the models with high MSE values rather that blindly combining the forecasts.

$$w_i = \frac{\prod_{j=0}^k MSE_j}{MSE_i \sum_{j=0}^k MSE_j} \quad w_i = \frac{\prod_{j=0}^k MSE_j}{MSE_i \sum_{j=0}^k MSE_j}$$

$$\dots\dots\dots(2)$$

$$where \sum_{i=1}^k w_i = 1$$

After assigning the weights we combine the forecasts to generate forecasts as follows,

$$\hat{y}_f = w_1 y_f^1 + w_2 y_f^2 + \dots + w_k y_f^k$$

$$\hat{y}_f = w_1 y_f^1 + w_2 y_f^2 + \dots + w_k y_f^k$$

$$\dots\dots\dots(3)$$

Where $f = 1, 2, \dots, T$ are number of forecasts

This combined forecast is expected to have better accuracy compared to individual forecasts. But this may not always be true. Hence as final step we select the forecast with least MSE value to be the final best

forecast. Further for comparison of the proposed algorithm and standard models we use most popular metrics such as Weighted Mean Average Percentage Error (WMAPE) and Root Mean Squared Error (RMSE) defined in equation 4 and equation 5 as follows

$$WMAPE = \frac{\sum_{k=1}^I (\hat{y}_k - test_k)}{\sum_{k=1}^I test_k} WMAPE = \frac{\sum_{k=1}^I (\hat{y}_k - test_k)}{\sum_{k=1}^I test_k}$$

..... (4)

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{N}} RMSE = \sqrt{\frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{N}}$$

..... (5)

V. CONCLUSION

In cloud environment, multiple virtual machine performance prediction is the task of predicting. Different resources by considering the differences from multiple task inferences based on the historical values to make effective and certainty judgmental decisions for the future values. There are different forecasting algorithms suitable for different workload type. Hence we proposed to combine different algorithms so that irrespective of the workload type it should be able to forecast the resource usage accurately. In this manuscript we have implemented the proposed ensemble of automatic algorithms and compared its accuracy with the individual base algorithm. The results show that the ensemble approach is able to forecast the vales for different workload types with improved accuracy. Further this model is general and can be applied to other than cloud dataset. The project can be extended to identify the required number of virtual machines in cloud environment in order to balance the workloads of various applications.

REFERENCES

- [2] Giuseppe Cicotti, Luigi Coppolino, Salvatore D'Antonio, Luigi Romano, "Big Data Analytics for QoS Prediction Through Probabilistic Model Checking", <http://arxiv.org/abs/1405.0327>, 2014. [accessed on 10th March, 2016].
- [3] Joseph Tsidulko, The 10 biggest cloud outages of 2014. <http://www.crn.com/slideshows/cloud/300075204/the-10-biggestcloud-outages-of-2014.htm/pgno/0/> 1. [accessed on 10th March, 2016].
- [4] S. Seneviratne, D C Levy, R Buyya, "A Taxonomy of Performance Prediction Systems in Parellel and Distributed Computing grids", CoRR, <http://arXiv.org/abs/1307.2380>. [accessed on 10th March, 2016]
- [5] "A little Book of R Time Series", Avril Coghlan, <https://media.readthedocs.org/pdf/a-little-book-of-r-for-timeseries/latest/a-little-book-of-r-for-time-series.pdf> [accessed on 10th March, 2016].
- [6] Zhijia Chen, Yuanchang Zhu, Yanqiang Di, Shaochong Feng, "SelfAdaptive Prediction of Cloud Resource Demands Using Ensemble Model and Subtractive-Fuzzy Clustering based Fuzzy Neural network", a research Article, Hindawi Publishing Corporation, Computational Intelligence and Neuroscience, Volume 2015, Article ID 919805, pp 114
- [7] Rongdong Hu, Jingfei Jiang, Guangming Liu, Lixin Wang, "Efficient Resources Provisioning Based on Load Forecasting in Cloud", a research Article, Hindawi Publishing Corporation, Volume 2014, Article ID 321231, <http://dx.doi.org/10.1155/2014/321231>
- [8] Yexi Jiang, Chang-Shing Perng et al, "Cloud Analytics for Capacity Planning and Instant VM Provisioning", IEEE Transactions on Network and Service Management, 2013.
- [9] Joao Loff et al, "Vadara: Predictive Elasticity for Cloud Applications", in proceedings of the 6th IEEE International Conference on Cloud Computing Technology and Science, December 2014, pp 541-546.
- [10] Qian Zhao, Guangsheng Feng, Ke Han, "A Mixed-Prediction based Method for Allocating Cloud Computing Resources", International Journal of Grid Distribution Computing, Volume 8, No. 2, 2015, pp 201212M. B. Kasmani, "A Socio-linguistic Study of Vowel Harmony in Persian (Different Age Groups Use of Vowel Harmony Perspective," *International Proceedings of Economics Development and Research*, ed. Chen D an, pp. 359-366, vol. 26, Singapore, 2011.
- [11] M Van Greunen, H A Engelbrecht, "Forecasting Methods for Cloud Hosted Resources, a Comparison", In proceedings of 11th International Conference on Network and Service Management, November 2015, pp 29-35.

- [12] Sajib Kundu et al, "Modeling Virtualized Applications using Machine Learning Techniques", ACM SIGPLAN Notices, 2012.
- [13] Rob J Hyndman, Yeasmin Khandakar, "Automatic Time Series Forecasting: The Forecasting Package for R", Journal of statistical software, Volume 27, issue 3, July.
- [14] J. Fritsh S, Gunther F, "Neuralnet: Training of neural networks", The R Journal, Volume 2/1, June 2010, pp 30-38.
- [15] Andreas Veit et al, "Household Electricity Demand Forecasting- Benchmarking State-of-the-Art Methods", In proceedings of 5th International conference on future energy systems, ACM, 2014, pp 233234.
- [16] T] Rob J Hyndman, YeasminKhandakar, "Automatic Time Series Forecasting: The Forecasting Package for R", Journal of statistical software, Volme 27, issue 3, July 2008*ransmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44-60.
- [17] Fritsh S, Gunther F, "Neuralnet: Training of neural networks", The R Journal, Volume 2/1, June 2010, pp 30-38.
- [18] Andreas Veit et al, "Household Electricity Demand Forecasting- Benchmarking State-of-the-Art Methods", In proceedings of 5th International conference on future energy systems, ACM, 2014, pp 233234.