

# Cloud Load Estimation with Deep Logarithmic Network for Workload and Time Series Optimization

**N. Bhalaji**

Associate Professor, Department of Information Technology, SSN College of Engineering, Kalavakkam, Tamil Nadu, India

**E-mail:** bhalajin@ssn.edu.in

## Abstract

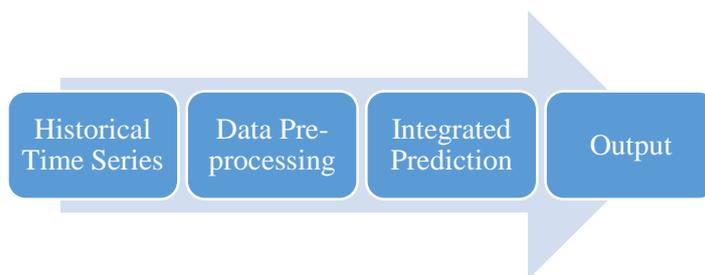
In recent days, we face workload and time series issue in cloud computing. This leads to wastage of network, computing and resources. To overcome this issue we have used integrated deep learning approach in our proposed work. Accurate prediction of workload and resource allocation with time series enhances the performance of the network. Initially the standard deviation is reduced by applying logarithmic operation and then powerful filters are adopted to remove the extreme points and noise interference. Further the time series is predicted by integrated deep learning method. This method accurately predicts the workload and sequence of resource along with time series. Then the obtained data is standardized by a Min-Max scalar and the quality of the network is preserved by incorporating network model. Finally our proposed method is compared with other currently used methods and the results are obtained.

**Keywords:** Workload, Time Series, Logarithmic Operation, Deep Learning, Min-Max Scalar

## 1. Introduction

Cloud computing has reached a massive growth and widely used in many companies and organization. It integrates CDC network, software, storage service and servers to produce network resource with pool of dynamic data storage and computing. The bandwidth is customized as per the user's requirement. Google, amazon, Facebook and Alibaba are few of

the classic providers of cloud computing. Interactive application is one of the successful applications of cloud computing [1, 2]. CDC is utilized by providers to maintain their quality of service, profit and customer satisfaction. Moreover this is cost effective. Proactive resource allocations are carried out to meet the resource availability and satisfy the norms of SLA. To achieve this we have to predict the workload accurately in CDC which is very difficult to accomplish. While execution the workload frequently changes and behaves dynamic and fluctuating. So the providers use payment model according to user's requirement and this model is termed as tiered business payment model [3, 4]. Here the user specifies the number of resources needed but some users are not aware of the model completely and couldn't give the actual count. This leads to wastage of time, resource, storage and service. This affects the revenue of the user. On the other hand some user specifies inadequate resource count and suffers with partially completed task. This leads CDC providers to loose users because of inadequate QOS and customer satisfaction [5, 6]. So the provider has to predict the workload and resource allocation along with the time series to meet the requirement.



**Figure 1.** Data Processing and BG-LSTM Structure

In traditional method they have concentrated on predicting workload and time series alone but this is insufficient to meet the needs of user and to provide QOS. So in our proposed work we have considered time series of both workload and resource allocation for better results. We have adopted Long Short-term memory (LSTM) and gated recurrent unit (GRU) to provide high quality results [7, 8]. The extended studies such as Bi-LSTM and Grid-LSTM provide some external changes to the network and used to predict pronoun of cross-language and

change in speech frequency of multi-channel. These methods are applied to frequently changing medium to capture their bi-directional characteristics [9]. The traditional method fails to satisfy the users need so in our proposed method we have integrated the applications of Bi-LSTM and Grid-LSTM with deep learning technique to produce a novel approach called Deep LSTM. The end result provides better resource allocation and workload prediction along with time slot [10-12].

## 2. Related work

The massive growth of CDC has attracted many users to utilize the service provided by CDC this increases responsibility of the providers to maintain good QOS. Many classical methods are implemented to predict the time series. Few of them are mention below

1. Autoregressive Integrated Moving Average Model (ARIMA)
2. Hidden Markov Model (HMM)
3. Support Vector Regression (SVR)
4. Back-Propagation Neural Network (BPNN)

In ARIMA the long term load prediction in developed cities are achieved by utilizing load variation. It is also used to predict highly fluctuating workload in public cloud. Since it is linear model it fails to predict the non-linear characteristics of workload [13, 14]. So they have moved to SVR which can predict the non-linear characteristics of time series data with multi-variant. Further HMM is used in virtual machine to predict the time dependency and tendency of the pattern of workload [15]. A combinational method of neural network and linear regression model produces error correction. Where the neural network is used as a sliding window and the linear regression model is used to measure the desirable resources but this method has many drawbacks and it is unsuccessful [16, 17]. Back propagation learning model incorporates Google cluster traces to evaluate the workload by using latency sensitivity.

Kalman filter and SVR acts as statistical learning module where the result appears with high accuracy than traditional methods [18-20]. When the time series increases it increase the prediction time and makes the system inefficient. A stochastic configuration network model is designed to predict the time series with the help of wavelet decomposition but this is only applicable for small dataset.

After the foot prints of artificial intelligence deep learning techniques have reached its peak and used widely. In recent years, prediction of time series based on deep learning method (Short-term) outperforms the traditional method (shallow learning method). The non-linear characteristics such as feature manifestation and hierarchical dissemination of the data are accurately evaluated by deep neural network [21, 22]. Deep belief network (DBN), stacked auto-encoder and long short term memory (LSTM) are some of the modern evolutions of deep learning techniques [20]. Out of this LSTM is capable of predicting workload in CDC. When LSTM performs independently it fails to meet the requirements of customers [23, 24]. So the researchers carried various experiment on LSTM to change its external structural characteristics. As result of these experiments we have obtained three successful methods namely Bi-LSTM, Conventional LSTM and Grid-LSTM. They differ by their connection lines which are used for the prediction process. The Bi-LSTM can be used to obtain the features of bio-informatics field. The digital sequence can be accurately sketched by using 2D Grid-LSTM. Which is used in real-time and its performance can be enhanced by applying recurrent connection along its dimensions. Integrating Bi-LSTM and Grid-LSTM we can accurately predict workload and resource time series of cloud data centres (CDC) [24, 25]. The above mentioned traditional methods have their own flaws and they are rectified by our proposed method.

### 3. Proposed Work

The framework of the proposed model comprises of two sub-divisions namely

#### a. LSTM

## b. Bi-LSTM

### 3.1 Long Short-Term Memory (LSTM)

The Long Short-term memory is used to predict time series in long term and it follows the pattern of past entry to predict the workload time series. By adding a memory cell, the LSTM improves the performance of classical RNN method. The input, status of the cell and output are determined by the gate unit present in each cell. The arithmetic expression of standard cell is given as

$$\begin{aligned}f_t &= \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \\i_t &= \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \\\tilde{c}_t &= \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}) \\c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\o_t &= \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \\h_t &= o_t \cdot \tanh(c_t)\end{aligned}\tag{1}$$

We have three gate units such as input, forget and output. From the above equation  $f$  denotes forget,  $i$  denotes input and  $o$  denotes output respectively. The input and states of the cell are denoted by  $x$  and  $c$  respectively. The bias and weight are the corresponding parameters of gate and cell state and they are represented as  $b$  and  $w$ . The recurrent information is represented as  $h$  and the time instant is represented as  $t$  and the past instant is represented as  $t-1$ . The activation of hyperbolic and sigmoid functions are represented as  $\tanh()$  and  $\sigma$  respectively. The multiplication function of vectors is defined by the operators. LSTM is capable of processing sequence irrespective of their length and the contextual information is processed by shorted long term memory. The major drawbacks of RNN are solved by LSTM.

It can be easily adapted to frequent variation in the system and the volatile time series is accurately predicted.

### 3.2 BG-LSTM

The accuracy is predicted by following three methods

1. The noise from the original data is removed by using SG filter. In addition the workload and resource are also considered as noise.
2. The scale of the original data is reduced by using Min-Max scalar and natural logarithm. This is called as pre-processing of data.
3. Finally the pre-processed data is trained and tested along time series by integrated Bi-LSTM and Grid-LSTM.

The pre-processed data act as input to the system. A sandwich model is designed where the Grid-LSTM is kept in between two Bi-LSTM layers. To obtain the output the end Bi-LSTM is connected to fully connected layer. Practically a single model cannot achieve accurate time series prediction. We need a combinational model to accurately capture the entire feature of time series. An improved bi-directional structure is constructed by Grid-LSTM to solve context sensitivity and gradient issue. To increase the accuracy of time series prediction we have to combine different models which can capture entire features of time series in different directions. Bi-LSTM and Grid-LSTM are combined to achieve better results. Bi-LSTM explicitly models the time series of current time slot and Grid-LSTM uses dimensional depth to model the time series. For accurate and deeper analysis of the data we have used two layers of Bi-LSTM.

The prediction model of BG-LSTM is explained in detail as

1. The historical data is obtained from the Google cluster and their timestamp information are analysed to determine the workload and number of resources

used. CPU and RAM are also considered to determine the time slot. Then this is considered as historical data.

2. Pre-processing of data is followed by three methods:
  - a. Natural logarithmic scale to reduce the raw scale values of workload and resource.
  - b. Using SG-filter the noise interference is removed.
  - c. Min-Max scalar is used to maintain the features of the data in same magnitude.
3. Finally the model is trained and tested by integrated Bi-LSTM and Grid-LSTM.

As discussed earlier the BG-LSTM model has a sandwich structure where the Grid-LSTM is kept in between two Bi-LSTM layers. The input of BG-LSTM is taken from the pre-processing data and the output is connected to a fully connected layer. Similar structures of LSTM are replaced by IL ( ).

$$O_t^L = \text{IL}(f_t^L, i_t^L, o_t^L, h_{t-1}^L, I_t)$$

$$O_t^{L+1} = \text{IL}(f_t^{L+1}, i_t^{L+1}, o_t^{L+1}, h_{t-1}^{L+1}, O_T^L)$$

$$O_{t+1}^{L+1} = \text{IL}(f_{t+1}^{L+1}, i_{t+1}^{L+1}, o_{t+1}^{L+1}, h_t^{L+1}, O_T^{L+1})$$

$$y_{t+1} = W_{h_y}^- O_{t+1}^{L+1} + W_{h_y}^+ O_{t+1}^{L+1} + b_y \quad (2)$$

From the above equation f represents forget, i represents input and o represents output respectively. The gate inputs are represented in terms of Forget, input and output. The input of BG-LSTM is represented by I and  $O_t^L$  represents the output of Bi-LSTM layer at L.  $O_{t+1}^{L+1}$  represents the output of Bi-LSTM layer at L+1.  $O_t^{L+1}$  represents the output of Grid-LSTM layer. The bias and weight are the corresponding parameters of gate and cell state and they are represented as b and w. The recurrent information is represented by h and the time instant is

represented as  $t$  and the past instant is represented as  $t-1$ . The sequence from  $t=1$  to  $T$  is represented by  $\rightarrow$  and the sequence from  $t=T$  to  $1$  is represented by  $\leftarrow$  and the output of BG-LSTM is represented by  $y_{t+1}$ .

**Table 1.** Performance of Workload Comparison with Various Filters

Method	RMSLE	MSE	R2
No filter	0.73	61631.37	0.92
Median filter	0.47	40782.85	0.94
Average filter	0.19	15988.28	0.98
<b>SG filter</b>	<b>0.16</b>	<b>13934.55</b>	<b>0.99</b>

**Table 2.** Performance of CPU Comparison with Various Filters

Method	RMSLE	MSE	R2
No filter	0.79	416.29	0.82
Median filter	0.58	284.69	0.90
Average filter	0.22	162.35	0.97
<b>SG filter</b>	<b>0.17</b>	<b>128.89</b>	<b>0.99</b>

**Table 3.** Performance of RAM Comparison with Various Filters

Method	RMSLE	MSE	R2
No filter	0.81	531.26	0.89
Median filter	0.56	302.92	0.93
Average filter	0.22	189.72	0.96

<b>SG filter</b>	<b>0.15</b>	<b>131.39</b>	<b>0.99</b>
------------------	-------------	---------------	-------------

#### 4. Result and Discussion

Here we have discussed the performance of proposed method with real time data.

##### 4.1 Pre-Processing of Data

The data are collected from Google cluster trace to evaluate the performance of proposed method. We have considered 12000 computers with workload of 25462157 task and 672003 jobs. Time slot is assigned for this work accordingly. It is taken for 29 days. The programs are in Linux. By analysing this timestamp information the workload and time series of resource are obtained. The obtained result is not accurate because it contains noise and extreme points. We have to remove it to obtain accurate data. The workload and resource are non-linear, non-stationary and changes frequently so it is very difficult to predict accurately. Natural logarithm is applied before smoothing. This reduces the magnitude of the workload and resource. This makes the task easier. When natural logarithm is applied to the original data it reduces the standard deviation. The processed data is connected to SG filter to remove noise interference. This filter is comparatively best among various filtering techniques such as RMSLE, MSE, median and average filter and R2.

The training and testing process is carried out by taking the Google dataset and the process is divided into three groups.

- a. Training (First 16 days)
- b. Verification (Middle 4 days)
- c. Testing (Last 9 days)

Then we reorganize the pre-processed data to preserve the originality of the data. Then the input is connected to the integrated model. It contains four groups and used for prediction.

- a. Bi-LSTM layer-2
- b. Grid-LSTM-1
- c. Dense-LSTM-1

Systematic investigation is performed multiple times with different combinations of BG-LSTM to select the best among them.

## 4.2 Prediction Result

To achieve better accuracy the BG-LSTM is tested on three variants namely

1. Original workload
2. CPU
3. RAM

The testing result of original workload, CPU and RAM gives higher accuracy with total workload along time series. To obtain precise output numerous simulations are carried out on random data of workload and resources. We have compared traditional methods (ARIMA and SVM) with our deep learning methods (LSTM, Bi-LSTM, Grid-LSTM, SG-LSTM, SG-Bi-LSTM and SG- Grid-LSTM) in terms of MSE, RMSLE and R2. These are the evaluation metrics used to evaluate the performance of the traditional and proposed methods. Since SG filter is used to remove the noise in our proposed work we have specified it along with each method. It is observed that better accuracy is acquired after the intrusion of SG filter. From RMSLE result it is observed that BG-LSTM provides higher accuracy value.

**Table 4.** Performance of BG-LSTM (SG-LSTM)

Method	RMSLE	MSE	R2
BG-LSTM (SG-LSTM)	0.16 (13934.55)	0.99 (73116.81)	0.23 (0.72)

From the comparative results it is observed that the BG-LSTM produces higher accuracy than that of SG-LSTM. BG-LSTM can capture some important features of time series such as time, frequency domain and bi-directional dependencies. BG-LSTM has higher modelling ability and fitting ability than the SG-LSTM. So BG-LSTM outperforms SG-LSTM. The comparative results of training loss prove that BG-LSTM has less training loss than other methods (SG-LSTM, SG-Bi LSTM and SG-Grid LSTM). When the epoch increases the training loss decreases. BG-LSTM is faster than other three methods. The combination method of BG-LSTM has higher modelling capacity than individual Bi-LSTM and Grid-LSTM. As discussed earlier the Bi-LSTM layer is capable of capturing explicit time series model along currently used time slot, bi-directional information and encoding property. The Grid-LSTM is capable of capturing time series using deeper dimensions. The output will give features of time and frequency domain. Then they are concatenated to produce modelling ability from LSTM. Thus the result proves that BG-LSTM is performing better than other methods.

## 5. Conclusion

A Cloud data centres (CDC) require highly accurate prediction model of workload and used resources to provide effective platform with better quality of service. This is very challenging task because of the continuously varying characteristic of workload and resource with the time series. Our proposed model have used integrated Bi-LSTM and Grid-LSTM layers for the prediction of workload and resource time series and SG-filter is incorporated to make the task easier by removing the noise interference in the system. Then BG-LSTM is implemented to analyse the characteristic of the workload and resource to improve the accuracy

of prediction in CDC. Finally it is applied to real time data set from Google cluster trace to prove that BG-LSTM is outperforms the traditional method.

In future, the issues like high dimensional input data and sparse data in a network can be rectified by applying our proposed method. By adding intelligent optimization to our proposed work the performance of the system gets enhanced and it can be applied for precise prediction in real-time application. Further the robustness of the system can be enhanced by performing numerous testing and training of various data types. It is proved that our proposed method is highly accurate, robust and effective one by outperforming the traditional methods.

## Reference

- [1] G. Luo and Q. Dong, "Progress Indication for Deep Learning Model Training: A Feasibility Demonstration", *IEEE Access*, vol.8, pp. 79811-79843, 2020.
- [2] Shakya, Subarna. "IoT based F-RAN Architecture using Cloud and Edge Detection System." *Journal of ISMAC* 3, no. 01 (2021): 31-39.
- [3] Sharma, Minakshi, Rajneesh Kumar, and Anurag Jain. "Load balancing in cloud computing environment: A broad perspective." In *Intelligent Data Communication Technologies and Internet of Things*, pp. 535-551. Springer, Singapore, 2021.
- [4] L. Zhang et al. "Real-Time Power System State Estimation and Forecasting via Deep Unrolled Neural Networks", *IEEE Transactions on Signal Processing*, vol.67, no.15, pp. 4069-4077, 2019.
- [5] Manoharan, J. Samuel. "A Novel User Layer Cloud Security Model based on Chaotic Arnold Transformation using Fingerprint Biometric Traits." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 36-51.
- [6] Hegde, Gayatri, and Madhuri Rao. "Smart Cloud: A Self-organizing Cloud." In *International Conference on Inventive Computation Technologies*, pp. 723-729. Springer, Cham, 2019.

- [7] K. Srivastava et al. "Data-Driven Day-Ahead PV Estimation using Auto Encoder-LSTM and Persistence Model", IEEE Transactions on Industry Applications, vol.56, no.6, pp. 7185-7192, 2020.
- [8] Raj, Jennifer S. "Security Enhanced Blockchain based Unmanned Aerial Vehicle Health Monitoring System." Journal of ISMAC 3, no. 02 (2021): 121-131.
- [9] Ghosh, Atonu, Debashis De, and Koushik Majumder. "A Systematic Review of Log-Based Cloud Forensics." Inventive Computation and Information Technologies (2021): 333-347.
- [10] A. Ahmed et al. "Machine Learning Methods for Spacecraft Telemetry Mining", IEEE Transactions on Aerospace and Electronic Systems, vol.55, no.4, pp. 1816-1827, 2019.
- [11] Smys, S., and Haoxiang Wang. "Security Enhancement in Smart Vehicle Using Blockchain-based Architectural Framework." Journal of Artificial Intelligence 3, no. 02 (2021): 90-100.
- [12] Mahaveerakannan, R., C. Suresh Gnana Dhas, and R. Rama Devi. "Cloud-Based Healthcare Portal in Virtual Private Cloud." In Inventive Communication and Computational Technologies, pp. 1071-1080. Springer, Singapore, 2020.
- [13] G. Zheng et al. "Prediction of Probability Density of Electric Vehicle Load Based on Deep Learning QRDCC Model", IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), pp. 1225-1229, 2019.
- [14] Sivaganesan, D. "A Data Driven Trust Mechanism Based on Blockchain in IoT Sensor Networks for Detection and Mitigation of Attacks." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 59-69.
- [15] Jayaraj, T., and J. Abdul Samath. "Cloud Based Heterogeneous Big Data Integration and Data Analysis for Business Intelligence." In International conference on Computer Networks, Big data and IoT, pp. 926-933. Springer, Cham, 2019.
- [16] C. Kang et al. "Clustering-Based Residential Baseline Estimation: A Probabilistic Perspective", IEEE Transactions on Smart Grid, vol.10, no.6, pp. 6014-6028, 2019.

- [17] Patil, Prachu J., Ritika V. Zalke, Kalyani R. Tumasare, Bhavana A. Shiwankar, Shivani R. Singh, and Shailesh Sakhare. "IoT Protocol for Accident Spotting with Medical Facility." *Journal of Artificial Intelligence* 3, no. 02 (2021): 140-150.
- [18] L. Pereira et al. "PB-NILM: Pinball Guided Deep Non-Intrusive Load Monitoring", *IEEE Access*, vol.8, pp. 48386-48398, 2020.
- [19] Bagde, Sejal, Pratiksha Ambade, Manasvi Batho, Piyush Duragkar, Prathmesh Dahikar, and Avinash Ikhari. "Internet of Things (IOT) Based Smart Switch." *Journal of IoT in Social, Mobile, Analytics, and Cloud* 3, no. 2 (2021): 149-162.
- [20] P. K.Ghosh et al. "Surrogate-Assisted Multi-Objective Probabilistic Optimal Power Flow for Distribution Network with Photovoltaic Generation and Electric Vehicles", *IEEE Access*, vol.9, pp. 34395-34414, 2021.
- [21] Madhura, S. "IoT Based Monitoring and Control System using Sensors." *Journal of IoT in Social, Mobile, Analytics, and Cloud* 3, no. 2: 111-120.
- [22] Y. Yuan et al. "WECC Composite Load Model Parameter Identification using Evolutionary Deep Reinforcement Learning", *IEEE Transactions on Smart Grid*, vol.11, no.6, pp. 5407-5417, 2020.
- [23] Hariharakrishnan, Jayaram, and N. Bhalaji. "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things." *Journal of ISMAC* 3, no. 02 (2021): 69-81.
- [24] D. Navarro et al. "Deep Learning-Based Magnetic Coupling Detection for Advanced Induction Heating Appliances", *IEEE Access*, vol. 7, pp. 181668-181677, 2019.
- [25] Suma, V. "Internet-of-Things (IoT) based Smart Agriculture in India-An Overview." *Journal of ISMAC* 3, no. 01 (2021): 1-15.

### Author's biography

**N. Bhalaji** has more than 15 years of teaching experience. He received his B.E. & M.E. degree in the discipline of Computer Science and Engineering and Ph.D. specializing in Trust Based Routing approach for MANETs from Anna University, Chennai. His current research interests

include Application of Trust over information and several communication domains namely Internet of Things and Blockchain Technologies. He is a recognised supervisor of Anna University and also Doctoral Committee member for VIT, SRM and Sathyabama University. He is a member of the board of studies in SRM Valliammai Engineering College and Vels University, Chennai.