

Adaptive Forecasting CPU usage of Virtual Machines

Sai Narsi Reddy Donthi Reddy
School of Electrical and
Computer Engineering
University of Missouri - Kansas City
Email: sdhy7@mail.umkc.edu

I. INTRODUCTION

In this paper we forecast CPU usage by five different virtual machines by using an adaptive forecasting technique based on neural networks.

In adaptive forecasting method, different forecasting techniques such as moving average (MA), weighted moving average (WMA), exponential smoothing (ES), exponential smoothing with trend (ESWT), window linear regression (WLR) and value from the previous day (Prev) are used as input to the temporal neural network. Some of these methods are explained in the section- II. Followed by adaptive forecasting method based on neural networks in section- III.

II. FORECASTING TECHNIQUES

A. Moving Average (MA)

In this technique, the forecasting is done by taking the simple average of last n data. i.e., to predict at time t we take the mean of values from $t - n$ to $t - 1$ as shown in the equal below.

$$F_t = \frac{\sum_{i=1}^n A_{(t-i)+1}}{n}, n \leq d \quad (1)$$

Where, F_t is the predicted value, A_t are the actual values and window size n should be less than or equal to the total size of the data d .

As this is a simple averaging of last n values, this method is computationally simple. However, if there is a lot of variation in the data, this method lags behind the data and produces lot of deviation.

In this paper to test the performance of the moving average method we choose window size of $\{5\}$.

B. Weighted Moving Average (WMA)

Similar to moving average technique, in this method we predict using the last n values. But, instead of given same weights to all the values we assign different weights based on the effect of value will be on the final prediction. In this method one thing to note is that the sum of all the weights should be equal to 1.

$$F_t = \sum_{i=1}^n W_i * A_{(t-i)+1}, n \leq d \quad (2)$$

Where, F_t is the predicted value, W_i are the weights whose some equals to 1 ($\sum_{i=1}^n W_i = 1$), A_t are the actual values and window size n should be less than or equal to the total size of the data d .

Performance of this method mainly depends on the weights and window parameters. In this paper, as CPU utilization data is collected every 5 min, if the utilization increases there is steep increase or vice verse. So, it is better to have more weight to the latest value in time series than to the old value. To test the performance we use same window sizes as above and for weight simple linear increments is used. i.e., if the window size is 5 then weights are assigned as $\{\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{4}{15}, \frac{5}{15}\}$.

C. Exponential Smoothing (ES)

In exponential smoothing the weight to the past predictions is given in exponentially decreasing manner. Unlike moving average method, where all the n values get equal weights, this method gives exponentially reducing weights and discounted the past data in a more gradual manner [1].

If a is the smoothing constant, F_t and A_t are predicted and actual values, then exponential smoothing is given as:

$$F_t = a * A_{t-1} + (1 - a) * F_{t-1} \quad (3)$$

To test the performance of this method we use a value of $\{0.8\}$. If a value is close to 1, then more weight is given to the latest actual value and less weight to the last predicted value. Exponential smoothing produces the results by taking interpolation between latest actual value and last predicted value [1].

D. Exponential Smoothing with Trend (EST)

The main disadvantage of exponential smoothing (ES) is that it does not follow the trend [2]. To overcome this double exponential smoothing or exponential smoothing with trend (EST) is proposed.

If FIT_t is the predicted value with trend, T_t is the trend for current predicting, then exponential smoothing with trend is given as:

$$FIT_t = F_t + T_t \quad (4)$$

$$F_t = a * A_{t-1} + (1 - a) * FIT_{t-1} \quad (5)$$

$$T_t = T_{t-1} + d(F_t - FIT_{t-1}) \quad (6)$$

Where, a and b are the prediction smoothing and trend smoothing constants.

For performance testing of exponential smoothing with trend same a value is used and $d = 0.5$.

E. Moving Window Regression (MWR)

Regression is statistical technique used to estimate the relation among variables [3]. This method follows the same technique as moving average, instead of taking the average of the values in given window we fit a polynomial line and predict the value.

To test the performance of the moving window regression, same window sizes as moving average are selected and 1st (linear regression).

F. Value from the previous day (Prev)

In this forecasting technique, simple windowed average of value around the same time t from the previous day is considered as the forecasted value to the present day time t . This method is useful specially in seasonal data.

III. ADAPTIVE FORECASTING METHOD

In this method all the mentioned forecasting techniques in section- II are considered as input X to a neural network. Then output P is predicted by multiplying X with weight matrix W as shown in equation-7.

$$P = X * W \quad (7)$$

Where X is a 1×6 matrix with all the forecasting techniques predictions and W is 6×1 weight matrix to combine all the forecasting techniques prediction to generate adaptive prediction P .

The weight matrix W is updated using an optimization technique called stochastic gradient descent (SGD). In SGD a gradiend of the error is used to update the network eights. In the case of temporal neural networks, weight matrix W is updated continuously.

$$E = Y - P \quad (8)$$

Where E is error for given target value Y and predicted value P .

Then the gradiend of the error w.r.t weight vector is calculated as follows:

$$\frac{dE}{dW} = E * X \quad (9)$$

However, as no input or output values are not normalized, the gradient in equation-9 will be really large and the weights will never converge and will be oscillating with large errors. So, to overcome this, gradient is normalized as follows:

$$\frac{dE}{dW} = \frac{E * X}{X * X^T} \quad (10)$$

Now weight matrix W is updated as follows,

$$W+ = l * \frac{dE}{dW} \quad (11)$$

Where, l is learning rate or step size. This functions same as the alpha value in exponential smoothing. Where we control the speed of variance in the forecating.

IV. EXPERIMENTS

A. Dataset

The CPU utilization data is collected from UMKC datacenter at an intervals of $5min$ for total of 501 virtual machines. For each virtual machine CPU utilization(MHz) is collect at $5min$ intervals for $6days$, generating average of 1442 values. For the evaluation purpose instead of considering all the virtual machines only 5 are considered randomly.

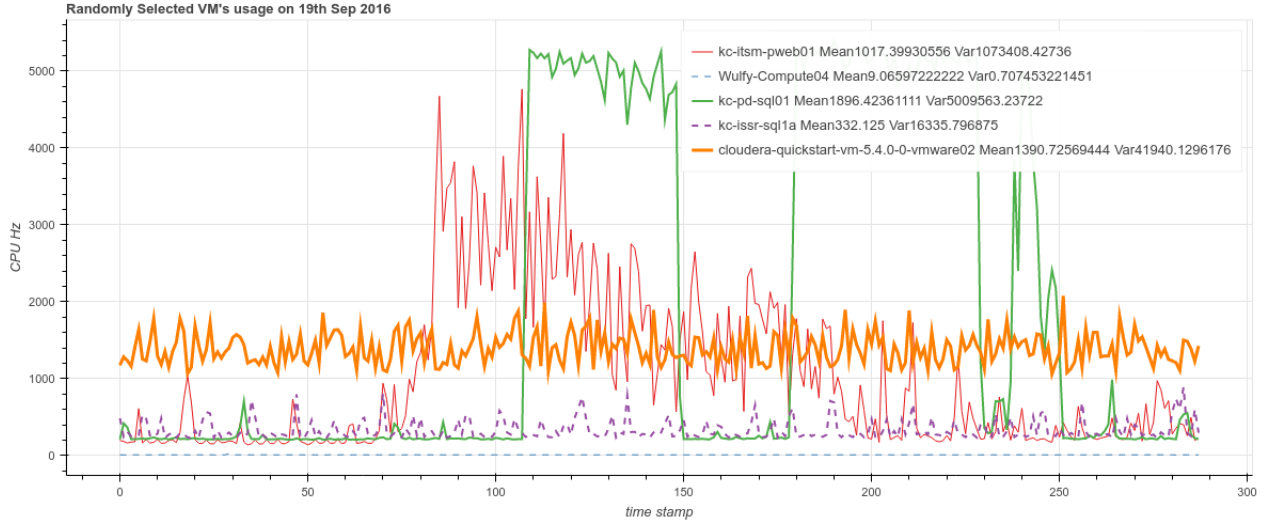


Fig. 1. Randomly selected virtual machines, their mean and variance

From Figure - 1, it can be seen that except virtual machine *Wulfy-Compute04* all the other virtual machines have very high variance, which results in constantly varying values and predicting these virtual machine CPU utilization will be really difficult.

B. Results

Figures - 2 to 7 show the adaptive forecasted results of five randomly selected virtual machines for the day of 19th Sep 2016. It can be seen that the negative slope effect caused by methods with trend such as ES, ESWT, and linear regression are mitigated by using adaptive forecasting technique with rectification.

This method as it can be clearly seen from the figure - 7 that the predicted signal tends to lag behind the actual value. This can be overcome by changing the step size or learning rate l in equation - 11.

The effect of l can be seen in figure - 2 and 3. If there is a sudden variation in the cpu utilization, the prediction is larger than the actual value, which can be seen in figure - 5. However, if the same variation in the future occur, the prediction value is not as large as the previous value because the weights are adopted to reduce effect. Which is not possible in simple forecasting techniques such as exponential smoothing and exponential smoothing with trend.

As the forecasting is performed through time, the weights are updated such that the more weight is given to the input which provide better prediction and less weights for the inputs with low importance in predicting the value. This phenomenon can be seen in figure - 8. It can be seen from the figure that more weight is given to weighted moving average (WMA) followed by previous values (Prev) and lowest weight is given to exponential smooth with trend.

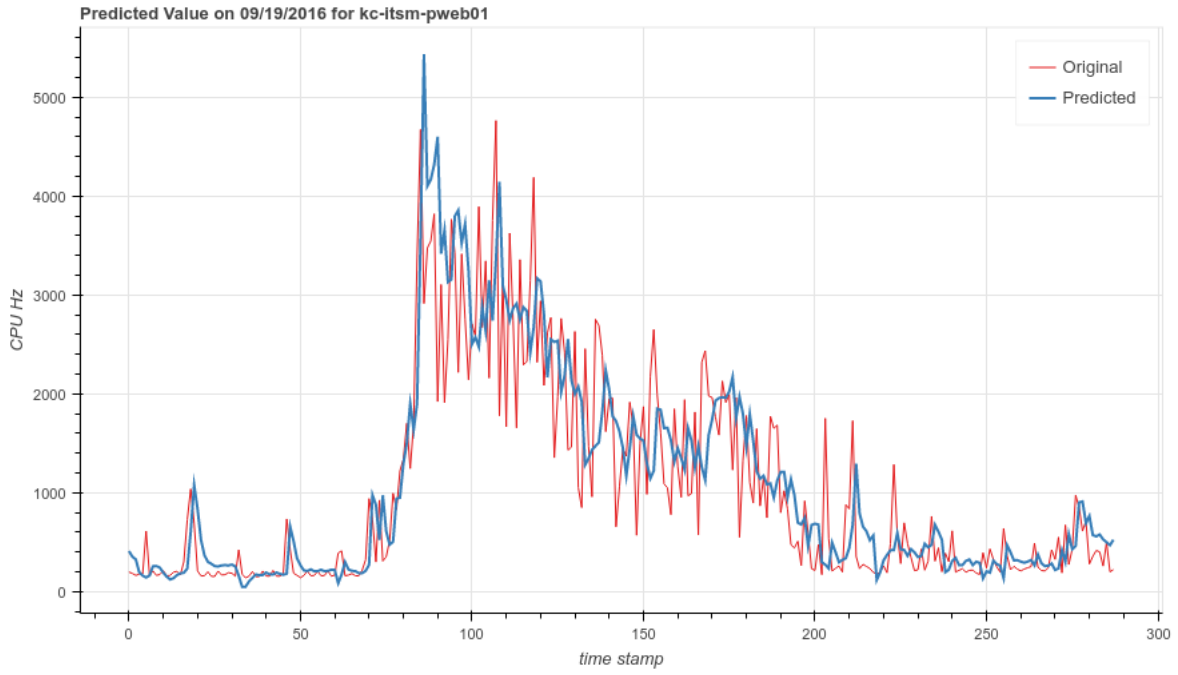


Fig. 2. Adaptive Forecasting result on randomly selected VM-1.

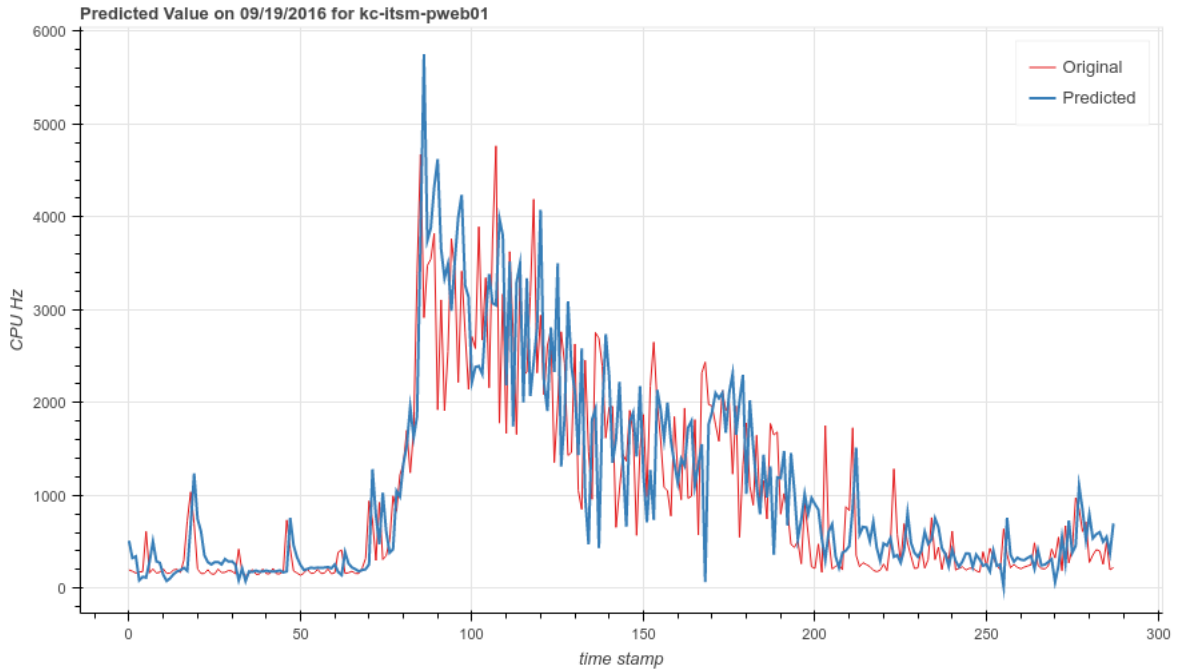


Fig. 3. Adaptive Forecasting result on randomly selected VM-1 with $l = 1$.

V. CONCLUSION

In paper a neural networks based adaptive forecasting technique is used to predict the CPU utilization of virtual machines in UMKC datacenter. In this method multiple forecasting techniques such as moving average, weighted moving average, exponential

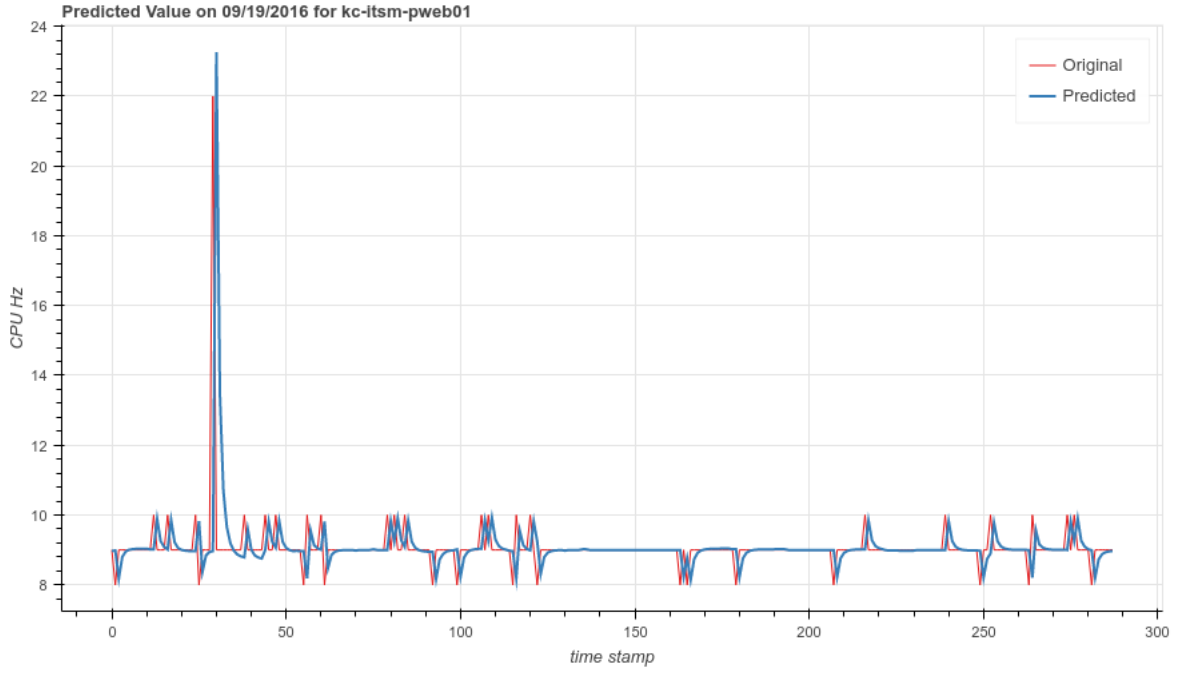


Fig. 4. Adaptive Forecasting result on randomly selected VM-2.

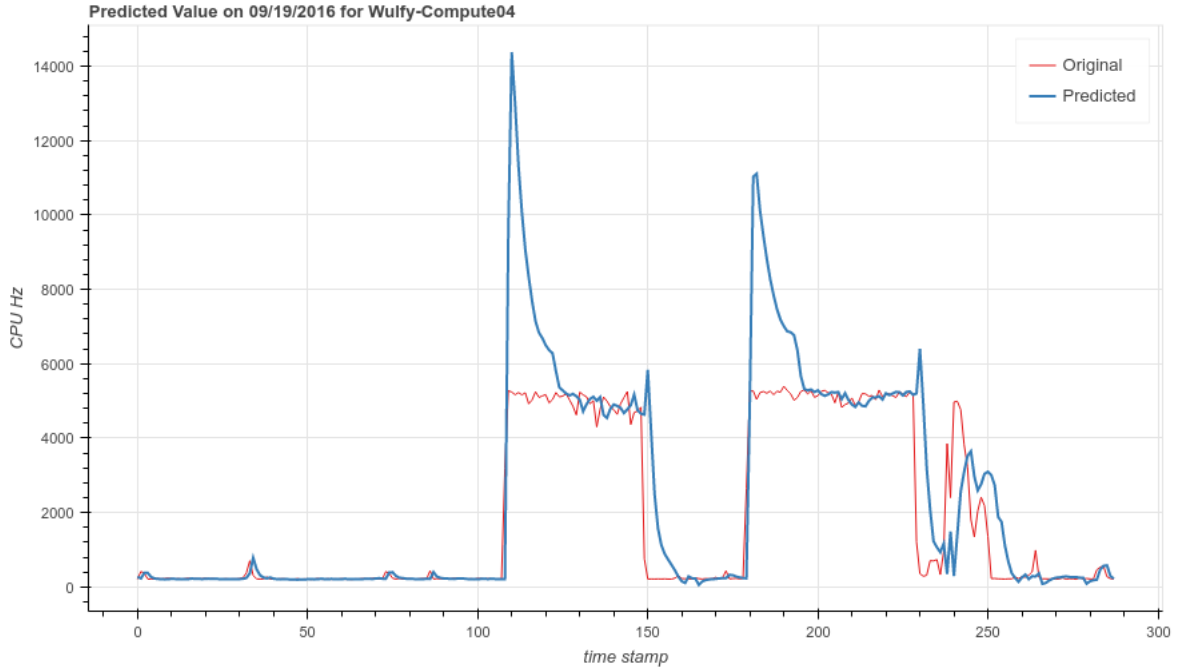


Fig. 5. Adaptive Forecasting result on randomly selected VM-3.

smoothing and exponential smoothing with trend etc., will be used as input to the temporal neural network and these inputs are combined to predict the CPU utilization. We also show how weights are updated for each input value through time.

In future, more tests will be conducted on this method and number of layers in neural network can be increased from single layer to multiple layers and also it is possible to increase number of input to the network by using different forecasting techniques or at least a window of previous values. The network in this paper is a simple linear unit. In future, nonlinearity can be added to this method to increase the performance of adaptive forecasting.

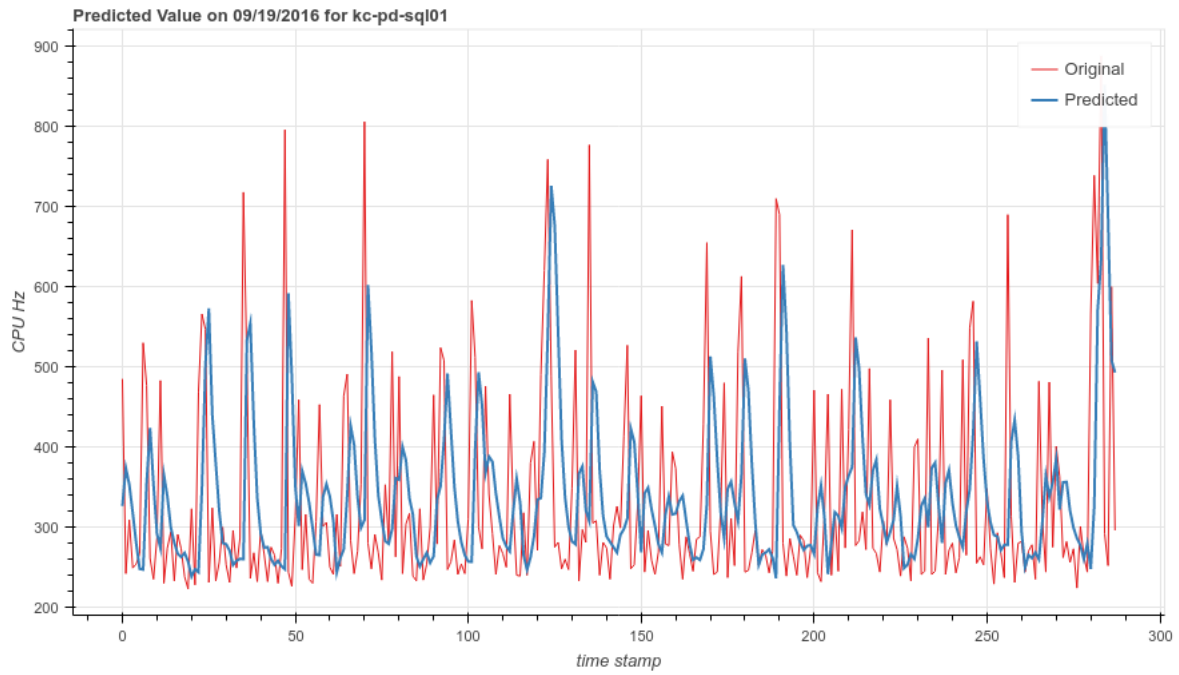


Fig. 6. Adaptive Forecasting result on randomly selected VM-4.

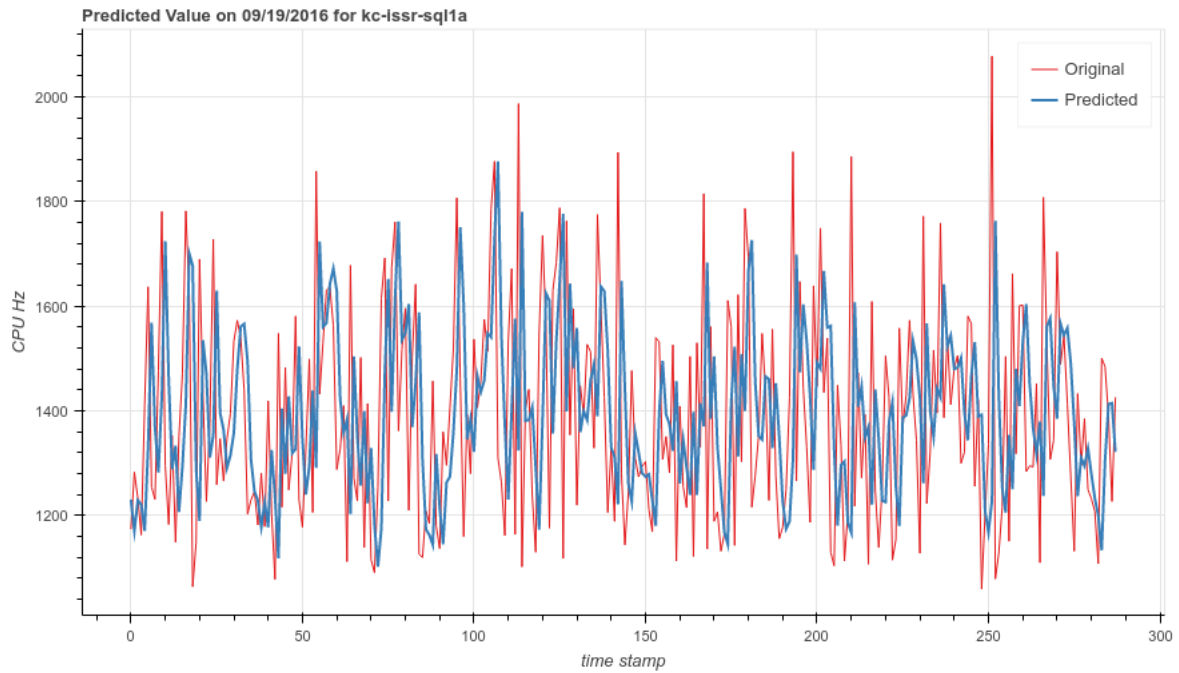


Fig. 7. Adaptive Forecasting result on randomly selected VM-5.

REFERENCES

- [1] Fuqua School of Business *Moving average and exponential smoothing models* <http://people.duke.edu/~rnau/411avg.htm#ses>.
- [2] NIST/SEMATECH *e-Handbook of Statistical Methods* <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc432.htm>
- [3] Wikipedia contributors *Regression analysis* https://en.wikipedia.org/w/index.php?title=Regression_analysis&oldid=738229734

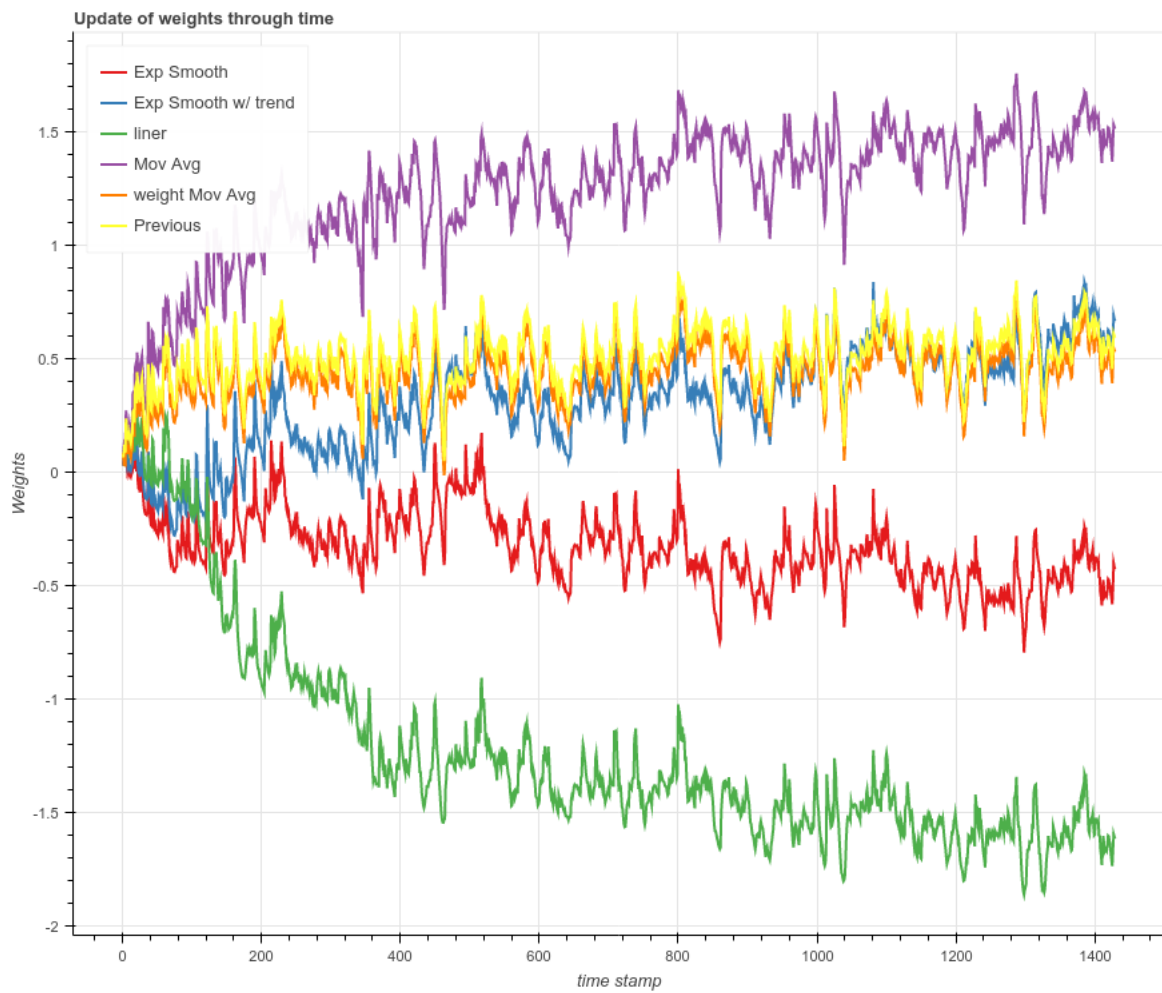


Fig. 8. The change in weights though time for VM-1.