Predictive Scaling Mechanism of Cloud Infrastructure using Machine Learning

Author – PUSHPALATHA K R, Hewlett Packard Inc, email : pushpalathakr9@gmail.com

Problem Statement

Major challenge is to decide when to scale and to decide the right number of resources to provision. There could be performance interference when behavior of one VM adversely affect the performance of another due to contention in the use of shared resources. The cost of a service cloud will be less if we lease less virtual resources from the cloud provider, but performance will be affected when the peak load occurs. Conversely, if more leased virtual resources lead to performance improvement leading to higher cost.

Solution

Predictive auto-scaling systems will generate a scaling decision based on the future forecast. Auto-scaling is the strategy that

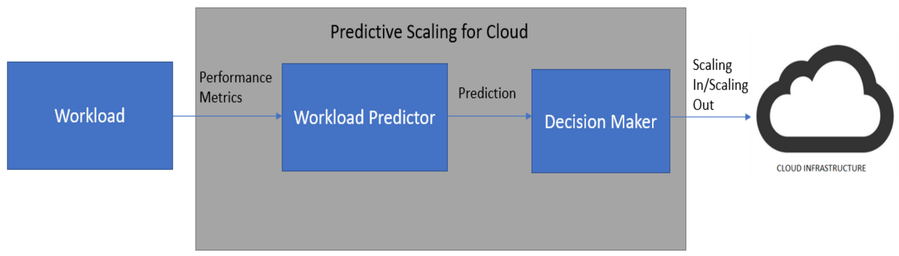

Fig 1: System Architecture

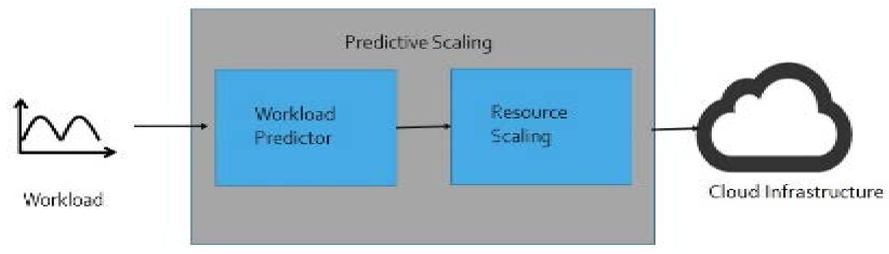

Fig 2: Predictive Scaling of Cloud Infrastructure

Methods

Support Vector Machine (SVM) Model: A training set D of n data points, D = {(xi, yi) |i = 1,...,n}, where x denotes an input vector and y is the output. In SVR training, our goal is to find the optimal hyperplane w. Time series prediction is used to estimate some future value based on current and past data samples.

Neural Network (NN) Model: Neural network classifiers are: feed-forward, back-propagation, time delay, and error correction neural network. Here, one output unit Y1 at the top layer (k=1) of each neural network for

regression problems can be considered.

Results

A predefined training dataset is defined for SVM and NN models. Both the models are evaluated against the testing dataset. 500 data points are used to try both models. 70% of the data points (approx. 350) of actual workload trace files are used as training dataset and the 30% (approx. 150) of the data points are used as testing dataset. Regression algorithms discover relations between features and use these relations to predict future value.

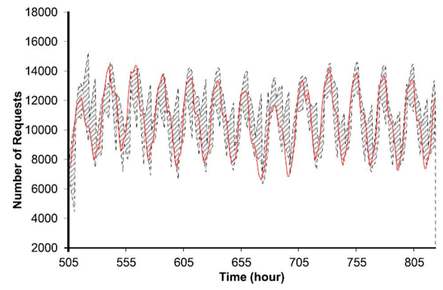

Fig 3: Workload of SVM Model Prediction with Actual and Predicted for a week

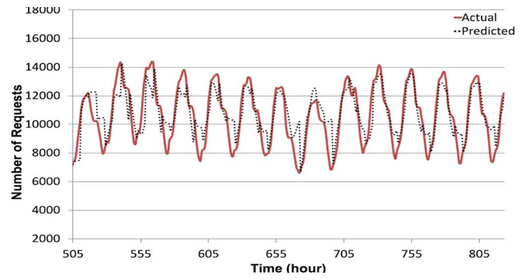

Fig 4: Workload of NN Model Prediction with Actual and Predicted

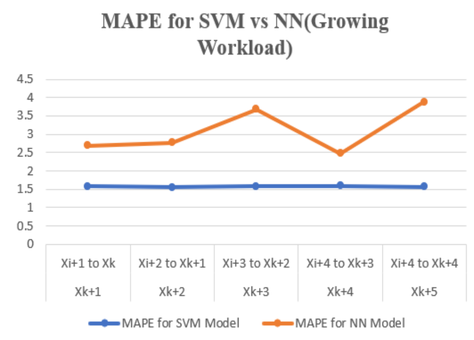

Fig 5: MAPE Plot for SVM vs NN for growing workload

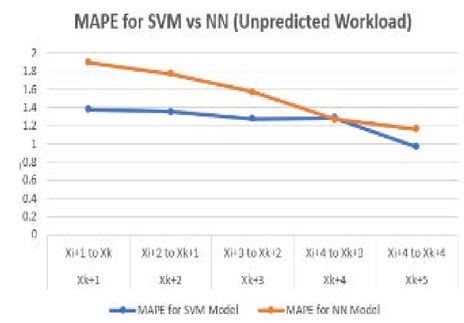

Fig 6: MAPE Plot for SVM vs NN for Unpredicted workload

Conclusion

SVM model is more accurate and better prediction model for growing and random-like(unpredicted) workload traffic pattern.

References

Support Vector Machines
http://www.statsoft.com/textbook/support-vector-machines
Empirical Risk and Structural Risk Minimization - https://mc.ai/empirical-structural-risk-minimization/