

Machine Learning Engineer Nanodegree

Capstone Proposal

Payam Mohajeri
18 December 2019

Domain Background

Virtualisation of application environments and containerisation of applications is a trending topic these days. Different solutions are provided to virtualise an operating system or containerise an application; and implement a cluster of nodes for managing or maintaining applications. One of the common topics within these clustering solutions is about distributing the load based on the available infrastructure resources, for example processing power and memory. For distributing the load, a component needs to keep monitoring the environment and provide the information to a scheduler to maintain the environment and organise future application deployments. The scheduler has an important role for improving and enhancing the environment or applications stability. One of the key metrics that scheduler has to consider is CPU utilisation. This project is going to look into providing a model for predicting the CPU utilisation.

Problem Statement

Improving the stability and performance of distributed applications across the globe is a challenging topic for enterprises. An application environment can get unstable when there is a heavy load on specific applications or nodes. Users often face high response time and low performance in case of any environment instability and this can have a huge impact on the branding or business of targeted enterprises and organisations. Today's competitive market requires organisations to provide stable and well performing application environment to their customers, but providing such stability based on a limited visibility on utilisation of infrastructure resources is very challenging. On this project I'm focusing on improving the visibility of system metrics by providing a model for predicting the CPU utilisation based on the previous infrastructure monitoring records. Besides improving the stability, saving energy and costs are also other use cases on such model.

Datasets and Inputs

Following provided datasets from Burn CPU Burn competitions on Kaggle will be used:
<https://www.kaggle.com/c/model-t4/data>

Solution Statement

The solution is to collect the system related metrics like the usage of memory, I/O, network traffic and train a model which can link the recorded data to the CPU utilisation and can provide an estimation in case the system behaviour is similar.

Benchmark Model

Information are all provided on <https://www.kaggle.com/c/model-t4/data>

Evaluation Metrics

Based on the information provided on <https://www.kaggle.com/c/model-t4/data> the evaluation metric is the Root Mean Square Error:

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{l}_i - l_i)^2}$$

Project Design

I intend to use what I have learned so far on this project, I will go through the discussions on Kaggle and implement a model on Jupiter notebook while deploying it on AWS. I'm interested to submit the results on Kaggle and compare / improve them based on the available rankings as well.