Survey of existing methods (Survey of literature)

Self-organizing maps (SOMs) is a well-known unsupervised learning algorithm based on a neural network approach using for clustering. This method can also be viewed as a constrained version of K-means clustering. SOM is efficient in handling large and high dimensional datasets. The method will reduce the difficulty of analyzing the high dimensional dataset by mapping down onto the two-dimensional coordinate system. SOM method has been used for anomaly detection and proved to be effective.

(\cite) used SOM for an anomaly detection in a server log data. The study presented an ability of the SOM method in detecting anomaly in the data, and also compared the results from the SOM method with an alarm threshold for each explanatory variable. The threshold for these univariate cases had been set to x+-2.28\*sd. Results showed that SOM method could detect more anomalies than all the obtained anomalies from every univariate tests.

Disadvantage

An initial weight vectors affect a performance of the SOM and lead to an unstable clustering results.

Require clean numeric data

The anomaly detection is not exactly what this thesis is mainly concerned.

Because the time series data in the analysis is rather fluctuate and not steady.

Detecting an anomaly in the data is one part of identify whether there are any changes in the data or not. In this study, the algorithm should be able to provide knowledge about the CPU utilization state and when changes. This objective leads to a Markov switching model which is later explained in more detail in sec 3.3.