

The normal behavior of the sensor network is characterized by series of observations over time. The problem of viewing anomalies can be viewed as finding deviations of a characteristic property in this data. There have been several approaches to this problem, including control charts, model-based methods, knowledge-based expert systems, pattern recognition and cluster analysis, hidden Markov models and neural networks. Most of them require prior knowledge about anomalous conditions/precise theoretical models of the system being monitored. But we would be interested in detecting any unacceptable/unseen change instead of looking for specific abnormal patterns.

Given this requirement, I think it would be interesting to investigate the effectiveness of negative selection algorithm for this problem. Briefly, the algorithm is as follows:

- 1) Define a set of observed (normal) values that would constitute the self-set S .
- 2) Generate a set of detectors D , such that each fails to match any element in S .
- 3) Monitor S for changes by continually matching the detectors in D against S .

There have been several improvements over this algorithm for effectiveness and efficiency. The representation of the elements of these sets depends on the range of values that the sensor measurements can take. We could use various normalization techniques to redistribute values in different ranges to a certain fixed range. But using any of the techniques like r -chunk rule or real-valued negative selection algorithm (which build upon the aforementioned naive algorithm) would require me to set parameters like r value or radius of hypersphere for self data. There are no predetermined ways to set these parameters, so it is one of the drawbacks of the technique.

As a starting point, I think I should implement the negative selection algorithm (r -chunk) and report some initial observations for the accelerometer data from Telegraph bridge. Please do let me know your views on this.

In parallel, I am also looking at the thesis on anomaly detection for time-series data. It talks about anomaly detection in multi-variate time series data, which might be of interest to us.