

Cyber Data Analytics - Assignment2

prerakmody

May 2019

1 Task1 - Familiarization Task

The BATADAL Dataset consists of 43 columns of raw data (X) and one column with the attack flag (Y). The different types of signals (few examples in fig. 1) that exist are :-

- Continuous signals - Eg : tank sensors for water level (L_T1 to L_T7)
- Binary Signals - Eg : Pump Actuator Status (S_PU1 to S_PU11) or ATT_FLAG (1's and 0's)

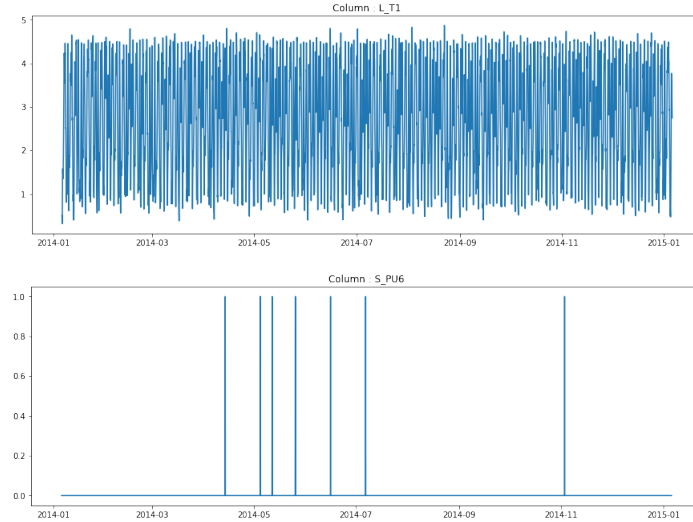


Figure 1: (Q1) Types of signals

We can notice in fig. 2 that there exists the presence of both highly positive and negative correlations between signals such as :-

- Strong Positive Correlation - F_PU2 and S_PU2 with P_J269 (i.e flow and status of pump2 is positively correlated with the pressure)
- Strong Negative Correlation - P_J269 with F_PU1 (i.e pressure is negatively correlated with flow in pump1)

We do a timeseries prediction task using linear regression as can be seen in fig. 3.

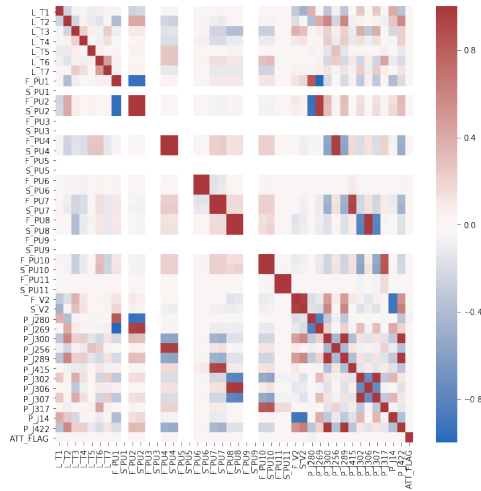


Figure 2: (Q1) Correlation between signals

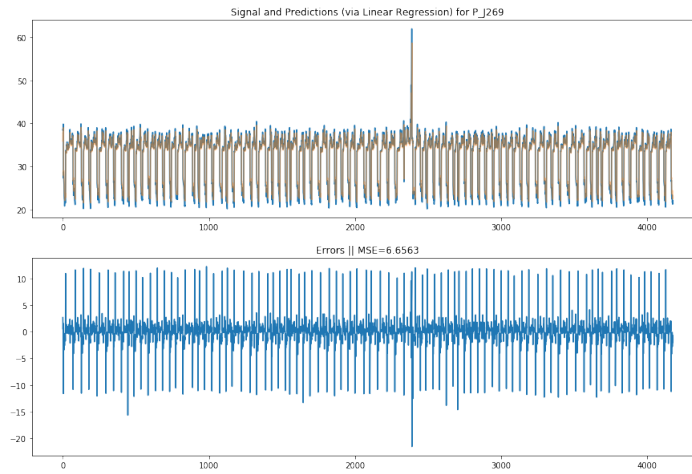


Figure 3: (Q1) Predicting time series values

- 2 Task2 - ARMA Task
- 3 Task3 - Discrete Models Task
- 4 Task4 - PCA Task
- 5 Task5 - Comparison Task