Anomaly Detection for Fleets of Gas Turbines

Nicholas Moehle

The goal of this project is to develop a data-driven fault detection and classification system for aeroderivative gas turbines. The approach an extension of multivariate statistical process control (multivariate SPC, or MSPC), which is heavily used in manufacturing and process industries. We hope to detect performance anomalies (faults) in a timely fashion (to prevent unexpected outages and further damage to turbines), and, if possible, provide information about the type of anomaly that occurred.

One of the major problems with the fault detection problem is the large amount of data. As a rough estimate, with 4500 turbines of the same model fielded at a time, and with data collected at 10 Hz, we might have:

$$4500 \text{ turbines} \times 50 \,\frac{\text{sensors}}{\text{turbine}} \times 10 \,\frac{\text{data points}}{\text{sensor} \cdot \text{second}} \times 8 \,\frac{\text{bytes}}{\text{data point}} \times 3 \cdot 10^7 \,\frac{\text{seconds}}{\text{year}} = 0.5 \,\frac{\text{petabytes}}{\text{year}}$$

With this amount of data produced per year, special attention must be paid to how these data are processed. Specifically, if each unit has a cumulative total of T data points per unit, and there are N units and k input channels and k output channels, we have the following approximate relation:

$$T >> N >> k$$

Any algorithm we produce must minimize computation time accordingly. In addition to detecting when an anomaly has occurred, we hope to learn something about the type on anomaly. First we divide anomalies into three types:

- *Point anomaly:* A sudden change in the performance of the turbine.
- *Drift anomaly:* A slow but consistent change in the performance of the turbine.
- *Unit anomaly:* A consistent difference in the performance of one turbine relative to the entire fleet.

We also hope to isolate the cause of these anomalies. For example, we may have a list of known point anomalies; we would like to reliably match these anomalies to a list of known anomaly hypotheses, or confidently reject all known anomaly hypotheses.

Due to the amount of available data, it is necessary to summarize the historical data in a compact form. We hope to do this using dynamic programming, with quadratic Bellman functions. this has the additional benefit that we can generalize some traditional results of SPC. Traditional SPC approaches, such as Shewhart charts, exponentially weighted moving average charts, and cumulative sum (CUSUM) charts can be cast as solutions to optimization problems, particularly as likelihood maximization problems. By looking at these optimization problems, we can generalize some of them to obtain techniques that are better suited to our problem.

The current optimization problem we are looking at has the following form

$$y_i(t) = B_i x_i(t) + a_i(t) + v_i(t)$$

where $x_i(t)$ and $y_i(t)$ are the (known) input and output of unit i at time t, $B_{i,t}$ is the static turbine map for turbine i, $a_i(t)$ is the bias (offset) for unit i at time t, and $v_i(t)$ is the residual for unit i at time t. We can devise three penalty terms:

- *size of residual:* $\|y_i(t) - B_i x(t) - a_i(t)\|$

- *bias drift:* $\|a_i(t+1) - a_i(t)\|^2$
- *deviation of model from fleet average:* $\|\bar{B} - B\|_{\mathrm{Fro}}^2$

We combine (the squares of) these penalty functions, and solve this optimization problem recursively. A particularly largle value in any of these terms roughly corresponds, respectively, to one of each of the above listed anomaly types.