Colloquium

KRISTIN BENNETT
Rensselaer Polytechnic Institute

SHAPE SIGNATURES: DATA MINING VIA GREY-BOX FEEDBACK CONTROL MODELS

Friday, November 30, 2018
3:00 p.m. in ES-143
(tea & coffee at 2:30 p.m. in ES-152)

ABSTRACT.
We present a new “grey-box” approach to data mining of streaming sensor data with applications in both smart manufacturing and systems biology. Our approach was first designed for tools run by control systems which execute recipe steps to produce semiconductor wafers. Multiple streaming sensors capture trace data to guide the control systems and detect anomalies. “White-box” models explicitly represent the underlying control (dynamical) system mathematically and are capable of capturing physical causal relationships, which can then be used in simulations to determine how the process will react to changes in control parameters. However, the underlying system is usually unknown, and even if it is known, explicitly modeling it is often intractable. Data-driven “black-box” approaches exist for anomaly detection, but they typically do not exploit the underlying control system. We propose a grey-box model, which possesses the strengths of both the white-box and black-box approaches. We use Bayesian regression to fit the parameters using a hypothesis set of damped linearly driven oscillators to form characteristic “shape signatures”. The Bayesian probabilistic model provides a natural anomaly score for each wafer, which captures poor control and strange shape signatures. We demonstrate how the anomaly scores can be used to monitor complex multi-step manufacturing processes to detect anomalies and changes and show how the shape signatures can provide insight into the underlying sources of process variation that are not readily apparent in the sensor data. We then demonstrate how the same shape signature approach can be applied to understanding natural control systems. Specifically, we utilize them to detect and characterize circadian rhythms in transcriptomic data from several species. By combining unsupervised learning and mathematical control models, the proposed “grey-box” methods produce valuable insights into regular and anomalous behaviors that are directly explainable in the original domain for both built and natural sensor data.