Big data analytics: improving batch operations in chemical plants

Martin Hollender, Moncef Chioua, ABB Corporate Research, Ladenburg, Germany

The FEE project on Early detection and decision support for critical situations in production environments

Big data technologies enable new possibilities to analyze historical data generated by process plants. One possible application is the development of new operator support systems, helping operators to run plants and a smooth and trouble-free way. The BMBF project FEE had the objective to develop such support functions based on big data analytics of historical plant data. The ABB Cooperate Research Center was working together with University of Kassel, Technical University Dresden and RapidMiner GmbH as development partners and BASF SE, INEOS in Cologne and PCK Raffinerie GmbH on specific application scenarios like event prediction, anomaly detection, and a batch profiling to investigate of the potential of Big Data technologies application context.

Application Scenario Batch Analysis

Jointly with the application partner BASF, batch process which showed unwanted behavior from time to time have been identified as interesting scenario within the project FEE. The goal for the FEE project was to use historical batch runs to identify potential problems in the currently running batch as early as possible. Data from more than 300 batch runs from a time period over two years was available for FEE project team.

Besides supervision of the batches in the control room, the operators have various other tasks inside the plant. One single operator can be responsible for several batches in parallel. An early warning from the FEE system that a batch might go into a wrong direction allows the operator to monitor this specific batch more closely and initiate countermeasures early so that the batch stays inside a good corridor.

One of the interesting learnings during the project was that the labeling of the batches into “good” and “bad” batches is not as straightforward as it first seemed. For example, a batch might have started to go into a wrong direction, but then operator intervened and made sure that the batch is produced smoothly. From the perspective of the plant, this is a “good” batch, whereas from the perspective of FEE it is a “bad” batch, because without operator intervention, the production problem would have occurred. However, there are various other reasons for operator interventions, which have nothing to do with a potential unwanted behavior.

During the production, it is often clear for the operators what is going on, but later on it is sometimes much less obvious why the trend curves look the way they are. Important operational context can get lost. BASF has installed additional tools that shall capture the semantics behind operator actions and why certain problems have occurred. In our case we found that many entries were missing, so that these tools didn't help us to label the batches better.

Another typical problem with batch production is that the trends of process variables from different batches need to be aligned. Although exact timestamps for start and end of the different batch phases exist, this is not sufficient for the comparison of two different batches with same recipe. A direct calculation of the Euclidean distance would result in a huge difference. A powerful technique to make two batches more comparable is called Dynamic Time Warping (DTW). DTW not only helps to align curves from different batch runs, it also computes a DTW distance metric on how similar two curves are. The DTW distance can be used to cluster the curves into groups of similar signals. Ideally, such clusters would separate between "good" and "bad" batches.

DTW clustering allowed us to identify a characteristic signature in one of the signals. If this signature is present, a production problem will follow with a very high probability later on.

As the clustering did not result in a sharp enough separation between "good" and "bad" batches, we have used multiway principal component analysis (MPCA) to classify the batches. This approach focuses on the process variables that most significantly contribute to the detected variability. With help of MPCA we were able to predict upcoming problems with a sufficient precision.