Manual: 4.1. Process Optimization

The purpose of machine learning is to develop a mathematical representation of the plant given the measurement data contained in the process control system. This representation necessarily needs to take into account the all-important factor of time as the plant has complex cause-and-effect relationships that must be represented in any model. These take place over multiple time scales. That is to say that the time from cause to effect is sometimes seconds, sometimes hours and sometimes days. Modeling effects at multiple scales is a complicating feature of the data.

The model should have the form that we can compute the state of the plant at a future time based on the state of the plant in the past and present. This sort of model can then be run cyclically so that we can compute the future state at any time in the future.

The state of the plant is the full collection of all tags that are important to the running of the plant. Typically there are several thousand tags in and around a plant that are important. These tags fall into three categories. First, we have the boundary conditions. These are tags over which the operators have no control. Examples include the weather and the quality of the raw materials. Second, we have the set-points. These are tags that are directly set by the operators in the control system and represent the ability to run the plant. Third, we have the monitors. These are all the other measurements that can be affected by operators (as they are not boundary conditions) but not directly (as they are not set-points) and so adjust themselves by virtue of the interconnected system that is the plant as the boundary conditions or set-points change over time. A typical example is any vibration measurement. This data is used by machine learning to obtain a model of the plant's process.

Having gotten the model, we want to use it to optimize the plant's performance. Here we need to agree on a precise definition of performance. It could be any numerical concept. Sometimes it is a physical quantity such as pollution (e.g. NOX, SOX) released, or an engineering quantity like overall efficiency of the whole plant, or a business quantity like profitability. We can compute this performance measure from the state of the plant at any point in time.

So now we have a well-defined optimization task: Find the values for the set-points such that the performance measure is a maximum taking into account that the boundary conditions are what they are. In addition to the natural boundary conditions (e.g. we cannot change weather) there could be other boundary conditions arising from safety protocols or other process limitations.

This is a complex, highly non-linear, multi-dimensional and constrained optimization problem that we solve via a method called simulated annealing. The nature of the task requires a so called heuristic optimization method as it is too complex for an exact solution. Simulated annealing has some unique features. It converges to the global optimum and can provide a sensible answer even if the time is limited.

The procedure in real-life practice is that we measure the state of the plant every so often (e.g. once per minute) by pulling the points from the OPC server and then update the model, find the optimal point, and report the actions to be performed on the set-points. This can be reported open-loop to the operators who then implement the action manually or closed-loop directly to the control system. As soon as something changes, e.g. the weather, the necessary corrective action is computed and reported. In this way, the plant is operated at the optimal point at all times.