Executive Briefing White Paper

Plant Performance Predictive Analytics

A Data Mining – Based Approach

ziconNET

Abstract

*The data mining buzzword has been floating around the process industries offices and control rooms. Yet, from buzzword to fad and from fad to necessity, data mining is finding its place in what is called **Level 3 domain of process systems**. Data mining has already established a presence in Level 4 Business Systems as means to do all sorts of business analytics and applications, and in Level 2 - Process Control Systems - in the form of one of its most popular predictive algorithm, Neural Networks, for advanced process control applications.*

Looking into wider uses of data mining in the process industries invites also use of a very appropriate term, Process or Manufacturing Plant Intelligence, which, like Business Intelligence, implies knowledge from plant history data by going beyond exploratory and multi-dimensional data structures for drilling down, slicing and dicing. It implies use of predictive models developed through data mining and for applications beyond inferential models. One popular application is failure prediction and root cause analysis. Yet, we look for data mining to serve needs, requirements and purposes that can be evolving and revolving. That is, establishing a platform that can be flexible for parallel initiatives or for redefining how to mine data and what to achieve.

This White Paper looks into how Data Mining can be applied for plant performance analytics by exploiting historical data and heat and material balances to predict stabilized conditions for a plan, or what the next steady state conditions will be. Once this prediction is made, the results can be further analyzed to see where for example may be significant deviations from design conditions, as this may imply a malfunctioning condition ranging from inefficiencies to leaks and from improper settings in the control system to loss of quality targets. Deviations from design conditions can be also explored further to assess performance of the entire plant or specific unit operations. The approach can be also applied for a complex consisted of various plants of similar design, which allows one to compare performance between different process plants.

Objective

Typically, plant operations consider a plant running well when it is running as close to its design conditions as possible. Or even better, when it is running at optimized conditions. Heat and material balances are used to make such benchmarking assessments, which are compared with ones derived from historical data. When looking into deviations, it is not always clear why these occur. Data Mining can identify such deviations and their causes, and as such, a useful objective is the ability to present plant performance on a comparative basis, against design conditions, thus, trying to draw conclusions about how a complex is running, has been running or where it is headed when it reaches stable condition.

Data Sets

To start with, we need design or optimum heat and material balances. The data set shown in the Appendix contains design heat and material balances from two plants. The data is transformed from a process simulation (using HYSYS) data sets for two identical plants (Plant A and B), as well as averages of stream measurements for each plant. The entire data set can serve as a basis to perform benchmarking and predictive analysis.

Benchmarking & Exploratory Data Mining

First, as a benchmarking exercise, various useful charts can be easily generated, as for example the one below – a scatterplot, showing how the temperature matrix of a plant (Actual Temperatures - AT) of both plants at various times fit against the equivalent design temperatures (Design Temperatures - DT) obtained from the set of design heat and material balances. The scatterplot below suggests that both plants A and B have been running fairly close to design with no significant deviations.

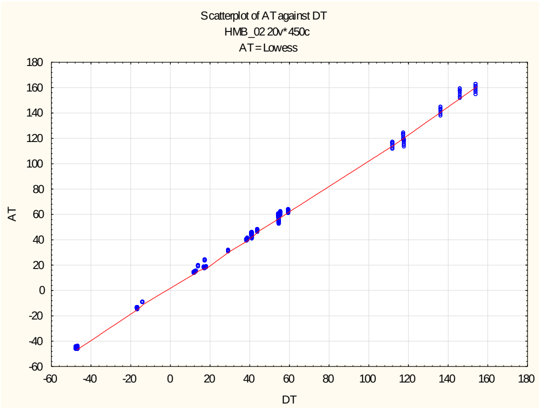

Figure 1 – Scatterplot of Actual versus Design Temperatures

In addition, any single stream variable, can be compared to design conditions by generating the line plot below, in this particular case, for stream S1 of Plant A.

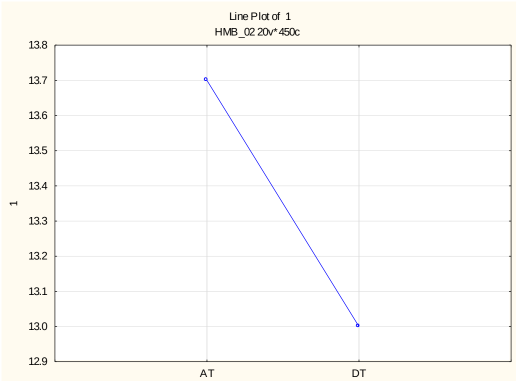

Figure 2 – Line Plot of Actual versus Design Temperature for Stream S1

Exploratory analysis can be further made with OLAP functionality. Starting from the existing structure of the data, a routine transformation can be performed on the data set to present stream variables as per the example below. It is pointed out again that validated and reconciled data can be included in additional to plain averages and aggregations from process historians.

Note that for the purpose of this white paper, Microsoft Excel 2010 provides sufficient functionality with Pivot Tables and Charts to generate this multidimensional presentation. Therefore, there is no need to invest into any costly OLAP software. For better or worse, Excel is a tool that practically everyone has, and especially, for the first time ever, Excel 2010 contains the required OLAP functionality.

Figure 3 – Exploratory Analysis of Actual and Design Heat & Material Balances

Predictive Analytics

The same data set can serve as bases for further Predictive Analytics, beyond routine forecasting. To explore this possibility following approach is followed.

First, the entire data of process streams (which can be measured or reconciled values) for various points in time is processed to identify relationships between process variables (Temperatures, Molar and Mass Flows, Pressures and Vapor Fractions). This is done by using a method known as **Feature Selection and**

Predictor Variable Screening

This method allows one, for any dependent variable, to identify a short list of good predictors from hundreds or even thousands of variables. It is also possible to rank the proposed predictors as well as the interactions between the predictors. This is a core function of data mining.

For our specific example, where we aim to predict a set of variables that define a stabilized condition, we define the plant's temperatures profile matrix as the target variables, because temperatures are the slowest response variables. The following relationships show that a predictive model, such as a Neural Network model, yields reasonable results for predicting the temperature profile matrix of the plant.

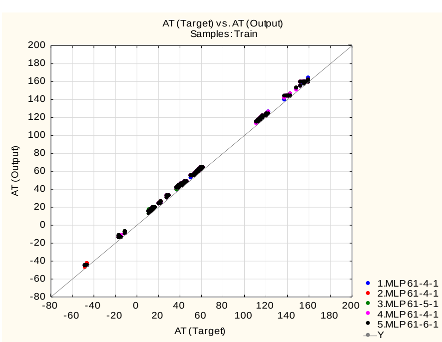

Figure 4 – Prediction of Plant Wide Temperatures at 100 % Plant Loads

With plants operating close to design conditions, one would not expect something very different, but when a similar model is applied to another set of data from plant A operated at 20% load, which means that the level of pressure and flow balance is substantially different between the 100% and 20% load of the plant, then again similarly reasonable relationships are obtained as shown below.

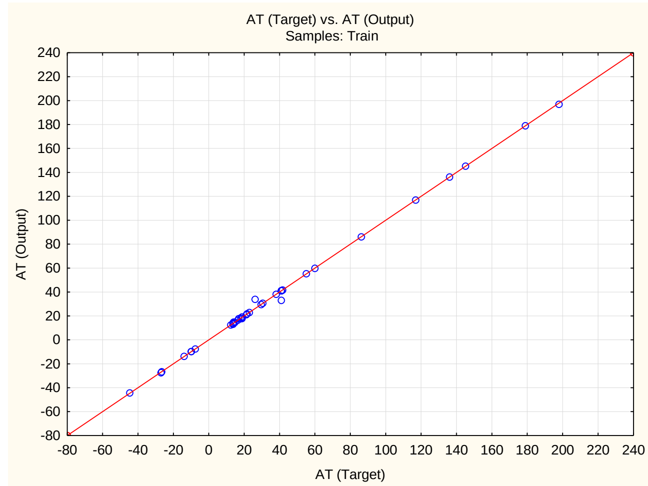

Figure 5 – Prediction of Plant Wide Temperatures including 20% Plant Load

Comparing Models

At this stage, the test is repeated in order to compare multiple models and derive the best predictive model, and for this purpose a standard workspace is applied as shown below. For this example, and from predictor correlation matrices, it is concluded that a Neural Network model gives the best results, although the Best-Subset and Stepwise ANCOVA and Standard Regression Trees (C & RT) give fairly close results.

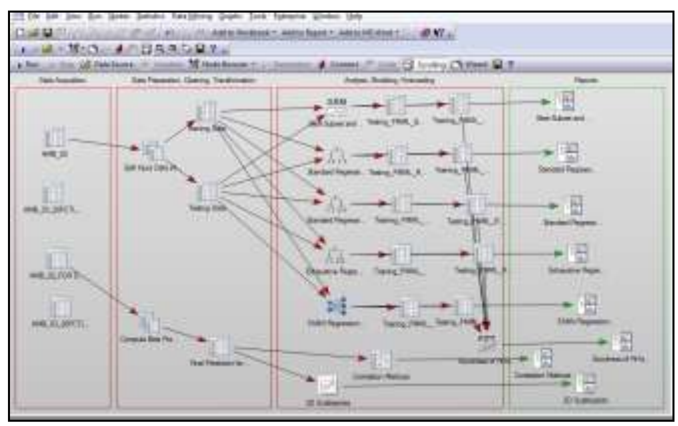

Figure 6 – Predictive Model Comparison Workspace

Once comparison of alternative models is achieved, the workspace shown above can include applying the best model to new data, as for example in the case of stabilizing a plant at 20% load and predicting the associated variables at this production level.

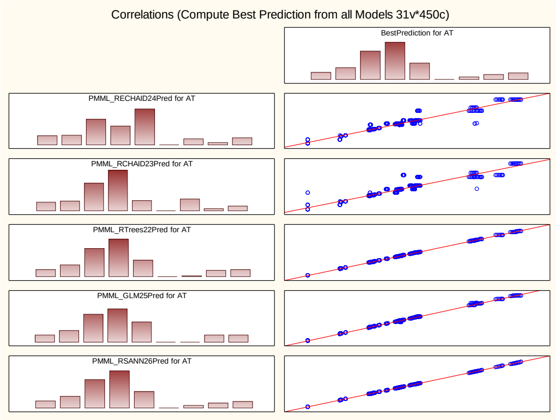

Figure 7 – Graphical Comparison of Predictive Models

Conclusion

Introducing Data Mining into process plant and manufacturing analytics makes benchmarking and plant performance much more productive. Presentation and visualization of benchmarking results with use of OLAP gives operations and management more insight into how their plants perform against design and/or optimized conditions. The workflows involved are very straight forward and can deal with massive data obtained from process historians. Introducing reconciled heat and material balance data makes such analyses even more useful.

Stelios Kentritas

Chemical Engineer and co-founder of ziconNET, executive consultant with >30 year career mainly in the process and IT industries and in the development and implementation of process systems, dynamic simulation, advanced process control, information management, manufacturing execution and supply chain management solutions, data warehousing and data mining systems, business intelligence and decision support systems, geographical information management, CAD/CAM/CAE, facilities management and GIS for land information and cadastre applications.

Author of a number of articles, white papers and extensive participation in forums as a speaker and visionary for information technology solutions and applications both in the process and business applications domains.

APPENDIX A: HEAT & MATERIAL BALANCE TEST DATA SET

	1 DATE	3 SNUMBER	4 DEVT	5 DEVP	6 DEVMR	7 DEVFR	8 DEVVR	9 AT	10 AP	11 AMR	12 AFR	13 AVR	14 DT
117	Mar/08	17	-2%	-4%	-7%	-6%	0.1%	61.2	17.9	158.2	5,325.6	97%	60.0
118	Mar/08	18	-4%	-6%	-5%	-5%	-0.4%	46.2	6.3	4.1	263.5	18%	44.5
119	Mar/08	19	-6%	0%	-7%	-6%	5.1%	58.0	64.8	154.1	5,062.0	64%	55.0
120	Mar/08	20	17%	-72%	-1972%	-2116%	0.0%	34.7	64.7	4,869.6	93,628.7	100%	41.7
121	Mar/08	21	-2%	-1%	-1%	-1%	0.0%	56.3	64.1	782.1	23,104.0	100%	55.2
122	Mar/08	22	-10%	-72%	-1972%	-2116%	0.0%	45.9	64.6	4,869.6	93,628.6	100%	41.7
123	Mar/08	23	35%	17%	52%	51%	0.0%	27.0	31.3	113.9	2,050.5	100%	41.7
124	Mar/08	24	5%	-3%	12%	18%	0.0%	17.6	64.1	4,869.3	93,624.1	100%	18.5
125	Mar/08	25	-5%	17%	52%	51%	0.0%	43.9	31.3	113.9	2,050.1	100%	41.7
126	Mar/08	26	27%	19%	52%	51%	0.0%	30.6	30.7	113.9	2,050.1	100%	41.7
127	Mar/08	27	-4%	-3%	-4%	-4%	0.0%	43.5	38.6	244.3	4,398.9	100%	41.7
128	Mar/08	28	32%	29%	52%	51%	0.0%	28.5	26.6	113.9	2,050.1	100%	41.7
129	Mar/08	29	-5%	16%	52%	51%	0.0%	43.9	31.5	113.9	2,050.5	100%	41.7
130	Mar/08	30	-5%	17%	52%	51%	0.0%	43.9	31.3	113.9	2,050.1	100%	41.7
131	Mar/08	31	-2%	-3%	45%	44%	0.0%	42.5	38.6	130.4	2,348.4	100%	41.7
132	Mar/08	32	-2%	31%	-4%	-4%	0.0%	42.5	26.1	244.3	4,398.5	100%	41.7
133	Mar/08	33	1%	-3%	-2%	-2%	0.0%	-45.6	29.4	4,725.3	85,099.9	100%	-46.2
134	Mar/08	34	3%	-1%	-6%	-6%	0.0%	-45.7	27.5	350.4	12,270.2	0%	-47.1
135	Mar/08	35	-36%	-4%	-6%	-6%	3.1%	18.4	26.4	350.4	12,270.1	48%	13.5
136	Mar/08	36	11%	-5%	-2%	-2%	0.3%	-14.3	63.6	5,651.4	116,728.0	90%	-16.0
137	Mar/08	37	-25%	-3%	-2%	-2%	0.0%	18.4	28.6	4,725.3	85,099.9	100%	14.6
138	Mar/08	38	-27%	-5%	-2%	-2%	0.0%	22.8	63.9	5,651.4	116,727.0	100%	18.0