Historical Process Data Based Energy Monitoring - Model Based Time-Series Segmentation To Determine Target Values

Janos J. Abonyia, Tibor T. Kulcsara, Miklos M. Balatonb, Laszlo L. Nagyb

a Department of Process Engineering, University of Pannonia, P.O. Box 158, H-8201 Hungary

b MOL Hungarian Oil and Gas Company, Szazhalombatta, Production Excellence

janos@abonyilab.com

Energy monitoring improves energy efficiency in process plants, by helping plant operators, engineers, and managers track actual and target energy. An energy monitoring system calculates actual energy use, estimates the energy needed when operating efficiently, tracks energy metrics, highlights performance issues, and prepares regular statements of energy use.

We developed a novel data-driven framework for energy monitoring. The developed on-line decision support system continuously evaluates the operation of the technology. Key Energy Indicators (KEI) including efficiency measures are calculated on-line based on empirical and first -principle models. Beside classical engineering approach correlation hunting techniques are applied to build these multivariate KEI soft sensors based on the selected process variables.

Energy monitoring is based on the comparison of KEIs and their target vales. In most of the cases these targets should depend on operating regimes. We developed advanced empirical modeling techniques to support on-line targeting. Comparison of the KEI-s at the same operating regimes allows the comparison of different operation strategies. Based on the extracted knowledge novel, realistic targets of KEI-s can be determined. To automatize this procedure we developed a goal-oriented time-series segmentation technique. With the proposed tool optimal target values for different operating regions can be determined.

A measure to evaluate the operation of the technology and comparison of performances of process operators has also been developed. Since the trends of these measures are the most important, the proposed framework utilizes a CUMSUM SQC chart for visualization and decision support.

The concept of the resulted historical data based energy monitoring system is demonstrated at Heavy Naphtha Hydrotreater and CCR Reform ing Units of MOL Hungarian Oil and Gas Company. The presented case study shows the applicability and efficiency of the proposed methodology.

* 1. Introduction

The purpose of monitoring and targeting (M&T) is to relate your energy consumption data to the weather, production figures or other measures in such a way that you get a better understanding of how energy is being used. In particular, it will identify if there are signs of avoidable waste or other opportunities to reduce consumption.

Data collection may be manual, automated, or a mixture of the two. Once an M&T scheme has been set up, its routine operation should be neither time-consuming nor complex. An M&T scheme will provide essential underpinning for your energy management activities, allowing you to:

1. Detect avoidable energy waste that might otherwise remain hidden. This is waste that occurs at random because of poor control, unexpected equipment faults or human error, and which can usually be put right quickly and cheaply (or, indeed, at no cost). Intercepting and rectifying such problems should more than cover the cost of operating the M&T scheme.
2. Quantify the savings achieved by any and all of your energy projects and campaigns in a manner that accounts fully for variations in weather, levels of production activity and other external factors. Many users cite this as the most valuable result of M&T.
3. Identify fruitful lines of investigation for energy surveys. Rather than starting a survey with no clear agenda, you can go prepared with specific questions to ask, prompted by observed erratic or unexpected patterns of consumption.
4. Provide feedback for staff awareness, improve budget setting and undertake benchmarking.

This guide presents M&T from two perspectives. One is routine use (on a weekly cycle, for example). Routine M&T as explained here is quick and simple and requires no particular expertise on the part of the user. The other perspective is target-setting and diagnosis, an aspect that will appeal to users who wish to analyze data in more depth. Although this aspect needs to be addressed when first setting up an M&T scheme, it then becomes optional once the system is up and running.

Even for the more advanced aspects, an appreciation of the basic physical principles behind the use of energy at your place of work and a basic grasp of math is all that is needed.

* 1. Energy monitoring concept

Methods for calculating expected consumption fall into two categories. There are those based on precedent (comparison with previous periods), and activity-based methods that relate expected consumption to its driving factors (weather, production throughput, mileage, etc.). For the sake of brevity we will refer to any procedure for calculating expected consumption as a ‘targeting model’. We use activity-based targeting models which calculate expected consumption by reference to its driving factors – the measurable things that cause consumption to vary.

The main concept of energy monitoring based on comparison of actual and targeted energy consumption. The difference between several energy monitoring methods is the determination of targeted consumption which can be prescribed or calculated. We developed energy monitoring technic for historical process data and we use linear regression model to produce a target energy use.

* + 1. Regression and deviance based classification

As it mentioned above we use linear regression model to calculate a possible energy consumption target. The linear model use a prediction equation (see eq 1) to perform a predicted value so the calculated model output is linear combination of process variables. In our case the is the energy consumption, is vector of process variables. To find predictor coefficient vector we use linear least squares method which find a solution for to minimize quadratic cost function.

|  |  |
| --- | --- |
|  | (1) |

The least squares problem has analytical solution for linear models and the solution for is the following:

|  |  |
| --- | --- |
|  | (2) |

Where is matrix where is the number of process variables and is the number of samples. is an vector of output variable (energy consumption).



Figure 1: Correlation diagram of measured and predicted energy consumption. The black line shows the ideal prediction; the dashed lines show Q levels based on standard deviation () of prediction error (). The lines belong to Q levels. We presumed normal distribution of prediction error.

The Figure 1 shows the correlation diagram of linear model. Due the behavior of Leas Squares Regression model this diagram helps us to qualify the energy efficiency of the technology. We presume that the range of model data is wide enough so covers operation rages with high and low energy consumption.

So when the predicted consumption is higher as the measured value the technology is considered to be efficient regards to historical data. The relation means that the technology could work with lower energy consumption.

* + 1. Confidence analysis

The uncertainty level of predicted value can be characterized based on variance analysis. The intent of our method is to produce a confidence range for each predicted point to find similar samples with prescribed probability.

Variance of dependent variable can be predicted by the variance and covariance of independent variables and

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (1) |

* 1. Regression based time series segmentation

The selection of appropriate data sequences can be considered as a time-series segmentation. Time-series segmentation is often used to extract internally homogeneous segments from a given time series e.g. to locate stable periods of time, to identify change points. A suitable segmentation algorithm has been presented in to extract homogeneous segments from historical or streaming data. We developed a fuzzy clustering algorithm to detect changes in the correlation structure of multivariate time-series in. Historical process data alone may not be sufficient for the monitoring of complex processes. We incorporated the first-principle model of the process into the segmentation algorithm in to use a model-based non-linear state-estimation algorithm to detect the changes in the correlation among the state-variables. For off-line purposes the bottom-up approach is widely followed, while for on-line application the so-called sliding window segmentation technique is applied. By combining and integrating recursive and dynamic principal component analysis (PCA) into time series segmentation techniques efficient multivariate segmentation method is obtained to detect homogenous operation ranges based on process data. These applications are mostly applicable in process monitoring since they can detect changes in the correlation structure of process data. However, now we are interested in extracting informative segments applicable for the energy monitoring.

The -segmentation of time series is a partition of to non - overlapping segments , such that , and . -segmentation splits to disjoint time intervals by segment boundaries, where . A time series is a finite set of -dimensional samples labelled by time points . A segment of is a set of consecutive time points .

The goal of segmentation is to find internally homogeneous, information rich segments from a given time series. To formalize this goal, a cost function describing the internal homogeneity of individual segments should be defined. This cost function is defined based on the distances between actual values of the time series and values given by a simple model of the segment. In univariate case this model can be constant or a linear function. For example in the sum of variances of the variables in the segment was defined as .

|  |  |
| --- | --- |
|  | (1) |
|  | (1) |

where the mean of the segment is considered as a simple model of the segment. Segmentation algorithms simultaneously determine the parameters of the models used to approximate the behavior of the system in the segments, and the borders of the segments by minimizing the sum of the costs of the individual segments:

|  |  |
| --- | --- |
|  | (1) |

The general cost function can be minimized by dynamic programming, which is unfortunately computationally intractable for many real data sets. Hence, usually one of the following heuristic, most common approaches are followed:

* Sliding window A segment is grown until it exceeds some error bound. The process repeats with the next data point not included in the newly approximated segment. For example a linear model is fitted on the observed period and the modeling error is analyzed.
* Top-down method The time series is recursively partitioned until some stopping criteria is met.
* Bottom-up method Starting from the finest possible approximation, segments are merged until some stopping criteria is met.
* Clustering based method time series segmentation may be viewed as clustering, but with a time-ordered structure. In a new fuzzy clustering algorithm has been proposed which can be effectively used to segment large, multivariate time series.

In data mining, the bottom-up algorithm has been used extensively to support a variety of time series data mining tasks for off-line analysis of process data. The algorithm begins creating a fine approximation of the time series, and iteratively merge the lowest cost pair of segments until a stopping criteria is met. When the pair of adjacent segments and are merged, the cost of merging the new segment with its right neighbor and the cost of merging the segment with its new larger neighbor must be calculated.

* 1. Results and discussion

A regressziós modell tanítása után a kapott paraméter vektorral becsült értékeket azok eredeti értékeinek függvényében ábrázolva egy redukált regressziós diagramot kapunk. Az ábrán feltüntettük a tökéletes illeszkedésű egyenes (fekete folytonos vonal), valamint az 1-2-3 határokat. A az hiba empirikus szórását jelöli. Ezzel sztenderd megbírhatósági határokat definiálhatunk a modellhez.

Az LKN módszer viselkedéséből adódóan a becslés egyfajta átlagként kezelhető, azaz az adott üzemi paraméterekhez a tanítóminta alapján egy átlagos energia felhasználást számol. A gyakorlatban ez azt jelenti, ha a becsült értékünk nagyobb, mint a hozzá tartozó mért érték, akkor a technológia üzemeltetése a vizsgált tartományhoz viszonyítva gazdaságos. Ha pedig a becsült érték alacsonyabb a mértnél, akkor a technológiát lehetne gazdaságosabban is üzemeltetni.

A regresszió konfidencia intervallum számítása alapján becslés jóságát minősíteni lehet. Variancia analízis alapján a becsült értékhez egy konfidencia szinthez () illeszkedő tartomány megadható. 0,98-as konfidencia szintet feltételezve a regresszió bemente alapján a becsült érték 0,98-as valószínűséggel a tartományon belül kell legyen.

Ha a tartomány a becsült térből áthelyezzük a mért értékek terébe, valamint illesztünk egy görbét , melynek független változója a mért érték, a függő változója a konfidencia intervallum sugara, akkor a következő tartományt kapjuk: . Belátható, hogy a konfidencia sávot a becsült tengelyen bárhol felvehetjük. Ennek értelmében bármely megadott eltéréshez adható egy intervallum, melyben a pontok valamilyen valószínűséggel ugyanahhoz a gazdaságossági szinthez tartoznak és ez a valószínűség a konfidencia szintje. A konfidencia sugara a modell tartományában változik, mégpedig a szélső tartományokban nagyobb. Ennek a oka, hogy a modell kevésbé megbízható a széleken mint a köztes intervallumban.

. 



