320883 Qualitative Trend Analysis With Shape Constrained Splines: Multivariate Extension and Validation With Full-Scale Data

Wednesday, November 6, 2013: 4:35 PM
Continental 7 (Hilton)

Kris Villez, Process Engineering, Eawag, Dübendorf, Switzerland, Christian M. Thuerlimann, Process Engineering, Eawag, Duebendorf, Switzerland and David J. Dürrenmatt, Rittmeyer AG, Baar, Switzerland

INTRODUCTION
Qualitative Trend Analysis (QTA) is a set of mathematical methods for segmentation of a time series in so called episodes. Such segmentation, i.e. a list of contiguous episodes, is referred to as a qualitative representation. The episodes themselves are characterized by a start time, end time and a primitive. This primitive is defined as a combination of particular sign for the measured variable, first derivative, and/or second derivative. Typically, one refers to these primitives with a unique character. In this work, the following primitives are considered:
A � convex antitonic (monotone decrease)
B � convex isotonic (monotone increase)
C � concave isotonic
D � concave antitonic
A popular use of QTA� is for fault diagnosis of batch and continuous processes. This is a result of (1) lacking detailed, principled knowledge about process dynamics under faulty scenarios and (2) availability of expert knowledge regarding anomalous conditions. However, many methods are based on heuristics and fail in terms of robustness to realistic noise levels [1]. For this reason, a method based on a combination of shaped constrained spline (SCS) fitting and the Branch-and-Bound (B&B) algorithm has recently been proposed with improved accuracy as a result [2-3]. Because computational efforts may be prohibitive for real-time applications, an approximating solution has been developed on the basis of a Hidden Markov Model (HMM) [4]. Both the SCS- and HMM-based method are suitable to univariate time series only. In this work, we present a multivariate extension of the SCS method for the first time.
Importantly, the newly developed method have only been tested thoroughly on a simulated data set. For this reason, the multivariate SCS-based method for QTA is also demonstrated on a full-scale data set obtained from the Winterthur wastewater treatment plant (WWTP). In contrast to the typical fault diagnosis application, it is applied here as a pure data mining method. More concretely, the method is applied to find the time of occurrence of inflection points in typical daily profiles of flow rate and oxygen measurements. It is hypothesized that tracking these particular times of occurrence in the long term can help in understanding seasonal, weekly and daily variations better.
METHOD AND INITIAL RESULTS
In the case of a univariate signal, one seeks to find the points in time at which the signal behaviour changes from one primitive to another. In Figure 1, the top panel shows the flow rate measurements during a single day. One can see that this typical profile roughly exhibits a *BCDA* (see above) sequence. Similarly, one can represent the oxygen measurements (bottom panel) as a *DABC* sequence. The SCS allows to find the times of the associated inflection points (*B* to *C* and *D* to *A* transition times) and the maximum (*C* to *D* transition) or minimum (*A* to *B* transition). This is based on a combination [2-3] of (1) Second Order Cone Programming (SOCP) for shape constrained spline fitting [5-6] and (2) the branch-and-bound algorithm [7-8].
The SCS method is extended as follows. First of all, the concept of an episode is generalized so capture the qualitative behaviour of two or more trends simultaneously. To this end, each episode is now characterized by as many primitives as there are multivariate signals. Because the transition times in the considered time series are not necessarily occurring simultaneously, this results in a larger number of episodes compared to the univariate case. The following sequence is realistic for the shown example and typical for the Winterthur plant:

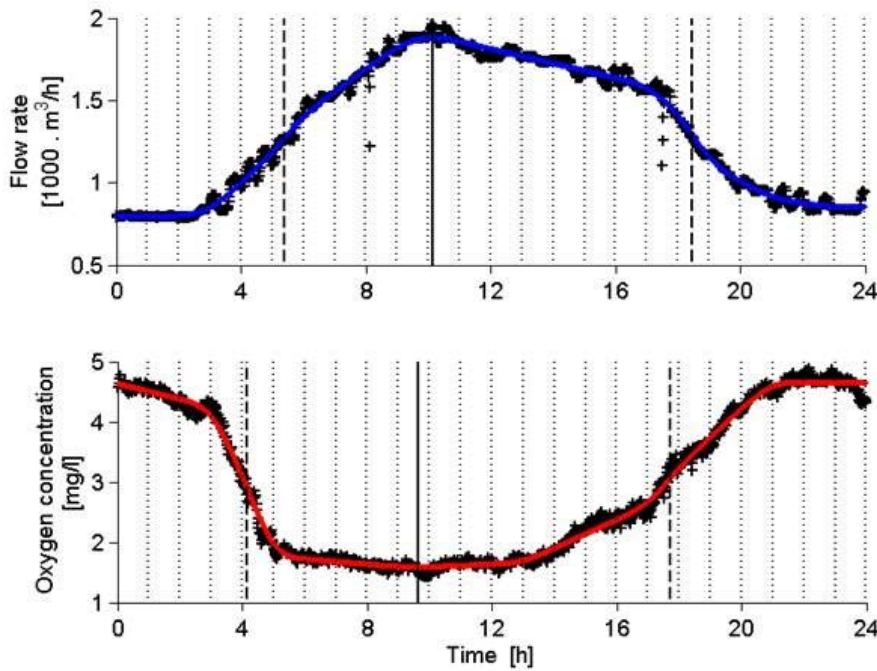| Episode index	1	2	3	4	5	6	7
Primitive 1	B	C	C	D	D	A	A
Primitive 2	D	D	A	A	B	B	C

Figure 1. Daily profiles (+) of flow rate measurements (top) and oxygen measurements (bottom). Vertical dotted lines (..) indicate the location of the spline knots. Vertical dashed lines (--) and full lines (-) indicate the identified inflection points, resp. maxima and minima, by means the Branch-and-Bound algorithm.

As expected, the resulting sequence is longer and has two primitives for each episode. In this case, 6 transition times have to be found.

The combined B&B/SOCP optimization scheme of [3] is easily extended based on the following elements. The objective function for the spline fitting (quadratic loss function) can be separated into a sum of individual objective functions for each of the considered time series:

$$J = \sum_i J_i$$

With J_i the quadratic loss for the ith time series:

$$J_i = || y_i - B_i . x_i ||_2$$

Where y_i is the column vector of measurements, x_i the corresponding spline coefficients and B_i the spline basis matrix. This spline basis is not necessarily the same for each time series, meaning that the spline order and knot placement can be set individually for each variable. In this work, a knot is placed every 15 minutes, corresponding to every 15th sample in a daily time series (1440 equally spaced measurements in total).

Similarly, each of the shape constraints associated with the given sequence of episodes (linear equality, linear inequality and second order cone constraints) are associated with one signal in the multivariate time series only. As a result, one can find the optimal spline coefficients associated with each series individually based on solving entirely separate Second Order Cone Programs (SOCPs). Solving the SOCPs assumes that one knows the transition times. For this reason, the branch-and-bound algorithm is also used here to optimize the transition times (as in the univariate case). As (1) the multivariate SCS fitting problem can be split into a number of univariate SCS fitting problems, and (2) upper ($J_{U,i}$) and lower ($J_{L,i}$) bounds have been proven for the univariate SCS fitting problem, one can write for the lower and upper bounds for the multivariate case that:

$$J_U = \sum_i J_{U,i}$$

$$J_L = \sum_i J_{L,i}$$

In other words, the upper (lower) bound for the multivariate problem is the sum of upper (lower) bounds for the individual univariate SCS fitting problems. This result makes it possible to apply the branch-and-bound algorithm in the same way as for the univariate case. Figure 1 shows the result obtained for a single day of operation. The generation of additional results for a long series of daily profiles is currently in progress.

CONCLUSIONS

A method for qualitative trend analysis of univariate signals has been extended for multivariate signals. Initial results obtained by the developed demonstrate the proper functioning of its implementation in Matlab for a single day multivariate signal. Detailed proofs and additional results for a longer time period are currently being generated.

REFERENCES

[1] Villez, K.; Rosén, C.; Anctil, F.; Duchesne, C.; Vanrolleghem, P.A. (2013). Qualitative Representation of Trends (QRT): Extended method for identification of consecutive inflection points. *Computers and Chemical Engineering,* 48, 187-199.

[2] Villez, K.; Rieger, L.; Keser, B. ; Venkatasubramanian, V. (2012). Probabilistic qualitative analysis for fault detection and identification of an on-line phosphate analyzer. International *Journal of Advances in Engineering Sciences and Applied Mathematics,* 4, 67-77.

[3] Villez, K.; Rengaswamy, R.; Venkatasubramanian, V. (2013). Generalized qualitative shape constrained spline fitting. *Computers and Chemical Engineering,* in review.

[4] Villez, K.; Rengaswamy, R. (2013). A generative approach to qualitative trend analysis forbatch process fault diagnosis. Accepted for oral presentation at the *European Control Conference,* Zurich, CH, Jul 17-19, 2013, Accepted for oral presentation.

[5] Nesterov, Y. *Squared functional systems and optimization problems.* In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (eds.) High performance optimization, applied optimization, vol. 33, pp. 405-440, Kluwer Academic Publishers, Dordrecht, 2000.

[6] Papp, D. *Optimization models for shape-constrained function estimation problems involving nonnegative polynomials and their restrictions.* M.Sc. thesis, Rutgers University, 2011.

[7] Mitten, L. G. (1970). Branch-and-bound methods: General formulation and properties. *Operations Research,* 18 , 24-34.

[8] Floudas, C. A., & Gounaris, C. E. (2009). A review of recent advances in global optimization. *Journal of Global Optimization,* 45 , 3-38.

Extended Abstract: File Not Uploaded