Computers and Chemical Engineering 58 (2013) 116–134

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Generalized shape constrained spline fitting for qualitative analysis of trends

Kris Villez[a,b,*], Venkat Venkatasubramanian[a,c], Raghunathan Rengaswamy[d]

[a] Laboratory for Intelligent Process Systems, School of Chemical Engineering, Purdue University, United States
[b] Spike, Department of Process Engineering, Eawag, Switzerland
[c] Department of Chemical Engineering, Columbia University, United States
[d] Department of Chemical Engineering, Indian Institute of Technology Madras, India

ARTICLE INFO

Article history:
Received 27 August 2012
Received in revised form 27 May 2013
Accepted 11 June 2013
Available online 4 July 2013

Keywords:
Data mining
Fault diagnosis
Qualitative trend analysis
Spline functions
Global optimization
Second order cone programming

ABSTRACT

In this work, we present a generalized method for analysis of data series based on shape constraint spline fitting which constitutes the first step toward a statistically optimal method for qualitative analysis of trends. The presented method is based on a branch-and-bound (B&B) algorithm which is applied for globally optimal fitting of a spline function subject to shape constraints. More specifically, the B&B algorithm searches for optimal argument values in which the sign of the fitted function and/or one or more of its derivatives change. We derive upper and lower bounding procedures for the B&B algorithm to efficiently converge to the global optimum. These bounds are based on existing solutions for shape constraint spline estimation via Second Order Cone Programs (SOCPs). The presented method is demonstrated with three different examples which are indicative of both the strengths and weaknesses of this method.

1. Introduction

This study was initiated in the context of qualitative analysis of trends. Such analysis concerns the abstraction of univariate or multivariate data series into qualitative descriptions. Most typically, qualitative analysis is focused on assessment of the sign of the first and/or second derivative of a time series. To do so, a data series is segmented into contiguous episodes within which the trends underlying to a data series are determined to have the same sign of the first derivative and/or second derivative. Such a segmentation is referred to as a qualitative representation (QR). We stress here that qualitative analysis of trends does not deal with the analysis of qualitative data which constitutes a particular branch in statistics of its own (Bryman & Burgess, 1993; Miles & Huberman, 1994). In what follows, we will use the term qualitative analysis for qualitative analysis of data series.

So far, qualitative analysis has primarily gained attention in the engineering literature. In this context, qualitative analysis enables to tie coarse-grained expert knowledge with automated, on-line data-based assessment tools. Typical engineering challenges attacked with this approach are process data mining (Stephanopoulos, Locher, Duff, Kamimura, & Stephanopoulos, 1997; Villez et al., 2007), reaction end-point detection (Villez, Rosén, Anctil, Duchesne, & Vanrolleghem, 2008, 2012) and, most popularly, process fault detection and identification (FDI) (Colomer, Melendez, & Gamero, 2002; Janusz & Venkatasubramanian, 1991; Maurya, Paritosh, Rengaswamy, & Venkatasubramanian, 2010; Maurya, Rengaswamy, & Venkatasubramanian, 2005; Rengaswamy & Venkatasubramanian, 1995; Rubio, Ruiz, & Mélendez, 2004; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Villez, Keser, & Rieger, 2009). The main advantage lies in the limited requirements for detailed process models or control design. Indeed, intuitive rule bases or tables can easily be generated and used to determine proper action once a QR of a time series is obtained (e.g., Villez et al., 2008).

A common characteristic of the existing methods for qualitative analysis is an extensive use of heuristics and/or tuning parameters and the absence of a well-defined global objective. For instance, the wavelet-based method developed in Bakshi and Stephanopoulos (1994) relies on a heuristic proposed by Witkin (1983) which determines the QR. The methods in Dash, Maurya, Venkatasubramanian, and Rengaswamy (2004) and Charbonnier, Garcia-Beltan, Cadet, and Gentil (2005) are based on piece-wise polynomial fitting. They make use of recursive schemes for on-line updating of the set of piece-wise polynomials. Because the optimal specification of interval endpoints for the piece-wise polynomials is NP-hard, locally optimal strategies based on lack-of-fit statistics are used.

* Corresponding author at: Spike, Department of Process Engineering, Eawag, Switzerland. Tel.: +41 58 765 5280; fax: +41 58 765 5389.
E-mail address: kris.villez@eawag.ch (K. Villez).

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.06.005

List of acronyms and symbols

a, b	left/right interval endpoint (scalar)
a_i, b_i	interval endpoints or knot arguments (scalar)
c_{ij}, x_{ij}, y_{ij}	element of matrix $C/X/Y$ on row i, column j (scalar)
c_i, x_i, y_i	ith element of vector $\mathbf{c}/\mathbf{x}/\mathbf{y}$ (scalar)
$\mathbf{c}$, $\mathbf{x}$, $\mathbf{y}$	column vectors
e	index of episodes (integer)
$f(t)$	univariate function in t
$f^{(k)}(t)$	kth derivative with respect to t of function f
$\mathbf{g}_{q+j,k}$	vector of linear coefficients to project the spline basis coefficients, $\mathbf{c}$, onto the piece-wise polynomial basis coefficients of degree k in interval $q+j$
h, i, j, q, r	index variables (integer)
k	order of monomial or derivative (integer)
l_e, u_e	lower and upper limit for right episode endpoint (scalar)
m	number of interval endpoints or knots (integer)
n	degree of polynomial or spline function (integer)
n_e	number of episodes (integer)
p_k	polynomial coefficient of order k (scalar)
$s_{e,k}$	sign of derivative of order k in episode e (+1, 0, or −1)
t	function argument (scalar)
$t_{e,1}$	left episode endpoint (scalar)
t_e, $t_{e,2}$, t_s	right episode endpoint (scalar)
B&B	Branch and Bound
$\mathbf{C}$, $\mathbf{X}$, $\mathbf{Y}$	matrices
DP	Dynamic Program
$F(\mathbf{y}, \mathbf{c})$	objective function in $\mathbf{y}$ and $\mathbf{c}$, convex in $\mathbf{c}$
F_L	Lower bound
F_{OPT}	Global optimum objective function
F_R	Reduced upper bound
F_U	Upper bound
MAP	Maximum A posteriori Probability
NLP	Non-Linear Program
QR	Qualitative Representation
QS	Qualitative Sequence
SCS	Shape Constrained Spline
SCSDP1	Shape Constrained Spline Diagnosis Procedure 1
SCSDP2	Shape Constrained Spline Diagnosis Procedure 2
SDP	Semi-Definite Program
SOC	Second Order Cone
SOCP	Second Order Cone Program
SSR	Sum of Squared Residuals
$\mathcal{A}$, $\mathcal{D}$	Sets of argument values
$\mathcal{C}_L$, $\mathcal{C}_U$	Set of spline coefficients for the lower/upper bound
$\mathcal{T}$	Set of feasible endpoints
λ_k	Regularization parameter for derivative or order k
σ	Noise standard deviation

Furthermore, it is indicated in Villez (2007) that these methods optimize the piece-wise polynomial representation to derive the QR rather than optimizing the QR itself. The methods in Cao and Rhinehart (1995) and Flehmig, Watzdorf, and Marquardt (1998) are based on filters set up for the assessment of the sign of a single derivative only. Joint analysis of the signs for more than one derivative has not been considered in this case. Multivariate qualitative analysis of trends, i.e., in which a multivariate data series is analyzed jointly, has only been considered so far in Flehmig and Marquardt (2006). Note that all existing methods consider a single explanatory variable only, which is time in all of the reported applications. We conclude this overview by stressing that the development and application of suboptimal methods can be a valid choice for on-line applications on grounds of computational speed.

Our primary motivation for this study is to provide a provably global optimal method for qualitative analysis. To this end, the problem of qualitative analysis is formulated for the first time as a shape constrained function fitting problem. Shape Constrained Spline (SCS) fitting is a non-parametric estimation problem with so called order restrictions on the spline function value and/or its derivatives. Alternative non-parametric models include kernel regression models and are almost exclusively related to monotonicity constraints (Delecroix, Simioni, & Thomas-Agnan, 1996; Dette, Neumeyer, & Pilz, 2006; Dette & Pilz, 2006; Hall, Huang, Gifford, & Gijbels, 2001; Holmes & Heard, 2003; Lavine & Mockus, 1995; Leitenstorfer & Tutz, 2007; Mammen, 1991; Meyer & Habtzghi, 2011; Robertson, Wright, & Dykstra, 1988). An exception is provided in Habtzghi and Datta (2012) in which other simple shapes are also considered, namely convex, monotone convex, or concave-convex. In Delecroix and Thomas-Agnan (2000) and Mammen, Marron, Turlach, and Wand (2001) both spline- and kernel-based approaches are discussed jointly.

A large fraction of the literature on SCS fitting deals with finding the optimal fit of positive, monotone, or convex spline functions, where optimality is defined as a least-squares or regularized objective function (Beliakov, 2000; Brunk, 1955; Dierckx, 1980; Fritsch, 1990; He & Shi, 1998; Kelly & Rice, 1990; Mammen & Thomas-Agnan, 1999; Meyer, 2008; Mukerjee, 1988; Ramsay, 1988, 1998; Tantiyaswasdikul & Woodroofe, 1994; Utreras, 1985; Wang & Li, 2008; Wegman & Wright, 1983; Wright & Wegman, 1980). More complex monotonic convex and monotonic concave shapes are applied by Elfving and Andersson (1988). In Hazelton and Turlach (2011), any shape constraint can be applied as long as it can be expressed in the form of a finite number of linear inequalities in the spline coefficients. To make this possible, the order of the spline is typically constrained to be equal or lower than two (linear) or three (quadratic). In this case, the feasible set of spline functions can be represented by polyhedral cones. Specific algorithms for such problems are provided in Dykstra (1983), Fraser and Massam (1989) and Meyer (1999). Alternatively, it is possible to formulate overly constrained solutions to ensure shape satisfaction of higher order splines. An approximate solution for monotone cubic splines in Meyer (2008) is based on solving first with a set of necessary but insufficient constraints which lead to satisfaction of the shape constraints in the knots of the splines. A subsequent interpolation procedure is used to obtain a spline function which satisfies the shape constraints over the whole function domain. In Turlach (2005), a more general method is proposed for splines constrained to any sort of shape. It is based on a two-step procedure in which violated constraints are identified and then mitigated by adding linear inequality and equality constraints. Both approaches are suboptimal because the resulting spline function is overly constrained. In contrast, Wang and Li (2008) recognize that sufficient and necessary conditions for monotone cubic splines can be formulated as SOC constraints. The more general idea that shape constraints on polynomials of any order can be represented as Semi-Definiteness constraints is presented in Nesterov (2000). SOC constraints are special instances of Semi-Definiteness constraints which result when the shape constrained polynomial is of third or lower order. As a result, spline functions up to 4th (cubic), 5th (quartic), and 6th (quintic) order with positive, monotone, and convex shape constraints, respectively, can be fitted efficiently with Interior Point optimization algorithms (Papp, 2011).

A few studies provide methods for more complex shape constraints (Hazelton & Turlach, 2011; Papp, 2011; Turlach, 2005). However, the provided methods assume that the intervals over which shape constraints are applied are known in advance. As a