Computers and Chemical Engineering 35 (2011) 63–70

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Identification of semi-parametric hybrid process models

Aidong Yang[a,b,*], Elaine Martin[a], Julian Morris[a]

[a] School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[b] Division of Civil, Chemical, and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, UK

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 3 February 2010
Accepted 1 May 2010
Available online 11 May 2010

Keywords:
Hybrid modelling
Model identification
Semi-parametric models

ABSTRACT

Hybrid models are mathematical models that comprise both mechanistic and black-box or data-driven components. Typically, the parameters in the mechanistic part of a hybrid model (if any) are assumed to be known. However in this research, a two-level approach is proposed for the identification of hybrid models where some parameters in the mechanistic part of the model are unknown. At the first level, the black-box component is identified using a regularization method with given values for the regularization and mechanistic parameters. At the second level, the regularization and mechanistic parameters are determined simultaneously and optimized according to a specific criterion placed on the predictive performance of the hybrid model. This approach is tested through the modelling of a toluene nitration process, where a support vector machine (SVM) model is used to represent the chemical kinetics, with the mass transfer-related mechanistic parameters being estimated simultaneously. The case study shows that good results can be obtained in terms of both the prediction of the process variables of interest and the estimates of the mechanistic parameters, when the measurement error in the training data is small whilst when the magnitude of the measurement error increases, the accuracy of the estimates of the mechanistic parameters decreases. However, the predictive performance of the resulting hybrid model in the latter case is still acceptable, and can be much better than that attained from the application of a pure black-box model under certain extrapolation conditions.

1. Introduction

In recent years, mathematical models for the characterization of chemical processes has become increasingly important for supporting various types of engineering tasks such as process design and process control. Depending on the information available, process modelling may be performed through a number of approaches including mechanistic, black-box (or data-driven), or hybrid modelling. By bringing together both existing mechanistic knowledge and data gathered from the process, a hybrid model that fuses both components has been shown, in a number of applications, to be advantageous when compared with a model formulated from either limited mechanistic knowledge or one constructed solely from the process data (Duarte, Saraivay, & Pantelides, 2004; Oliveira, 2004; Psichogios & Ungar, 1992; Thompson & Kramer, 1994). The advantages of hybrid models have motivated a number of applications, such as the modelling of batch polymerization reactors (Tian, Zhang, & Morris, 2001), fermentation processes (Saraceno, Curcio, Calabro, & Iorio, in press; Wang, Chen, Liu, & Pan, 2010) and boilers (Rusinowski & Stanek, 2010). Besides, Teixeira et al. (2007) discussed the general role of hybrid modelling in the combination of systems biology and process engineering.

When modelling a chemical process, the black-box model or the black-box component of a hybrid model will usually have the characteristics of a "universal approximator," i.e. one that is capable of approximating any arbitrary function. Within this group of models, artificial neural networks (e.g. Montague & Morris, 1994; Psichogios & Ungar, 1992; Thompson & Kramer, 1994) have most often been considered. More generally, a black-box model will belong to the family of non-parametric models (Eubank, 1988; Green & Silverman, 1994; Hastie & Tibshirani, 1990). When identifying a non-parametric model, regularization is often applied to address the issue of over-fitting, i.e. a regularization parameter is utilised as a weighting factor for the penalty term in the criterion for training. Thus, the development of a non-parametric model under regularization typically includes two tasks: the selection of an appropriate value for the *regularization parameter*, and the estimation of the *parameters of the non-parametric model* according to the selected training criterion. A detailed discussion on this topic can be found in the non-parametric regression literature (e.g. Geman, Bienenstock, & Doursat, 1992; Green & Silverman, 1994).

* Corresponding author. Current address: Faculty of Engineering and Physical Sciences (J2), University of Surrey, Guildford, GU2 7XH, UK. Tel.: +44 1483 686577; fax: +44 1483 686581.
E-mail address: a.yang@surrey.ac.uk (A. Yang).

0098-1354/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2010.05.002

With respect to the mechanistic component of a hybrid model, the most frequently utilised first principles knowledge are the conservation laws. These materialize in the mass/energy balance equations being incorporated into the model. In contrast, the mechanistic knowledge often missing from a model relates to the so-called constitutive relationships, i.e. those relationships that define the rates of (bio)-chemical reactions and transport phenomena, or those relationships that model the physical properties. In reported hybrid modelling studies (references cited above), black-box models are frequently adopted for approximating the unknown constitutive relationships, whilst the rest of the mechanistic knowledge is assumed to be available. The systematic methods for identifying this type of models can be found in, for example, Kahrs and Marquardt (2008). However, in reality a constitutive relationship may be of a known mechanistic form but has unknown parameters that require to be estimated. This relationship, and those described by non-parametric, block-box models, may require to be combined to give an overarching model. Such a situation has received little (if any) consideration to date in the process modelling literature.[1] Lima, Saraiva, and GEPSI-PSE Group (2007) reported a framework for establishing semi-mechanistic models by adding empirical elements into a mechanistic model which itself may have unknown parameters. These empirical elements, selected out of an "extension set", tend to be relatively simple, parametric expressions, therefore representing a class of models different from that addressed in this work.

A hybrid model that combines both a parametric and a non-parametric component can be termed a semi-parametric model. Semi-parametric models have been studied in detail in statistics but primarily in the context of data-driven modelling (cf. Haerdle, Mueller, Sperlich, & Werwatz, 2004; Ruppert, Wand, & Carroll, 2003). In contrast, the semi-parametric models considered in this paper are the result of hybrid modelling, where the model structures are typically derived from mechanistic knowledge and hence fail to conform to the typical semi-parametric model forms that have been studied in statistics, such as the generalized partial linear additive model (Haerdle et al., 2004).

Investigating regularization methods for ill-posed problems, Weese and co-workers (Roths, Marth, Weese, & Honerkamp, 2001; Weese, 1993) studied a type of model identification problems, where both an unknown function f and a number of unknown parameters a_i were to be estimated. More specifically, the model to be identified assumes the following form:

$$g(t) = K(f)(t) + \sum_i a_i h_i(t), \tag{1}$$

where K is a nonlinear operator of f, h_i is a function, t is time. Both K and h_i are known from theory. Regularization was introduced in the identification process; the regularization parameter was determined by means of optimization whilst the parameters of the finite-dimensional approximator of f as well as the "mechanistic" parameters a_i were identified simultaneously.

The focus of this paper is the identification of semi-parametric hybrid models of chemical processes which are generally different from those represented by Eq. (1). The problem is formulated in Section 2 whilst in Sections 3 and 4, a two-level identification approach and its implementation are described, respectively. This approach is different from the one by Weese and co-workers and allows to incorporate established black-box modelling algorithms into an optimization framework without changes to these algorithms. An application of the proposed approach is reported in Sections 5 and 6 with conclusions being presented in Section 7.

[1] This paper addresses those cases where a black-box model is developed for the modelling of a constitutive relationship. However, the proposed solution approach may also be applicable to other types of hybrid models that involve unknown mechanistic physical parameters.

2. Problem formulation

Consider the following functional relationship as a model, or part of a model, that characterizes a chemical process system:

$$y = f(x, v, p), \tag{2}$$

$$v = h(x), \tag{3}$$

where x and y are vectors of the independent and dependent process variables, respectively; v is a vector of process quantities which are a function of x; and p is a vector of constant mechanistic parameters. When the defined process is in a transient state, x, y, and v may vary with time.

This paper considers hybrid modelling scenarios that satisfy the following assumptions:

(a) The form of the function f is known as a consequence of underlying mechanistic knowledge.
(b) The form of the function h is unknown, hence a black-box model requires to be developed to approximate its form.
(c) p is unknown and requires to be estimated.

Furthermore it is assumed that the structure of f and/or the measurements of y and x are such that, according to Eq. (2), v can be computed analytically or numerically at the sampling points of x and y for a given estimate of p:

$$\tilde{v} := f^{-1}(\tilde{x}, \tilde{y}, \hat{p}). \tag{4}$$

It is noted that noisy measurements can pose problems for the computation of $\tilde{v}$. This issue is discussed in the case study in Section 6.

Based on these assumptions, the task of hybrid modelling, as studied in this paper, can be stated as follows: for a given set of measurements, x and y, an estimate of p and an approximation of h is obtained such that the resulting model has acceptable capability in terms of predicting the behaviour of the chemical process being modelled.

3. A two-level solution approach

To identify the hybrid models described above, a two-level approach, which is an adaptation of the framework for identifying non-parametric (black-box) models with regularization, is developed. Details of the regularized identification of the black-box models are first presented. Extensions to the approach are then proposed to address the estimation of the mechanistic parameters.

3.1. Regularized identification of black-box models

Following the notation defined above, a black-box model is considered:

$$\hat{v} = \hat{h}(x, \theta), \tag{5}$$

where $\hat{h}$ is an estimate of h (defined in Eq. (3)) which represents the form of the black-box model; θ is the vector of parameters of the black-box model. Under a regularization framework, estimation of the parameters, θ, is achieved through the minimization of a general function that takes the form:

$$I_1 = \sum_{i=1}^{M} c(\tilde{v}_i, \hat{v}_i) + \lambda R. \tag{6}$$