2016 Indian Control Conference (ICC)
Indian Institute of Technology Hyderabad
January 4-6, 2016. Hyderabad, India

Use Genetic Programming for Selecting Predictor Variables and Modeling in Process Identification

Devendra Verma, Purva Goel, Veena Patil-Shinde*, Sanjeev S. Tambe
Chemical Engineering and Process Development Division, CSIR-National Chemical Laboratory, Pune, India.
devendra@iitk.ac.in, ps.goel@ncl.res.in, veenaashinde@gmail.com, ss.tambe@ncl.res.in

***Abstract*—Availability of an accurate and robust dynamic model is essential for implementing the model dependent process control. When first principles based modeling becomes difficult, tedious and/or costly, a dynamic model in the black-box form is obtained (process identification) by using the measured input-output process data. Such a dynamic model frequently contains a number of time delayed inputs and outputs as predictor variables. The determination of the specific predictor variables is usually done via a trial and error approach that requires an extensive computational effort. The computational intelligence (CI) based data-driven modeling technique, namely, *genetic programming* (GP) can search and optimize both the structure and parameters of a linear/nonlinear dynamic process model.It is also capable of choosing those predictor variables that significantly influence the model output.Thus usage of GP for process identification helps in avoiding the extensive time and efforts involved in the selection of the time delayed input-output variables. This advantageous GP feature has been illustrated in this study by conducting process identification of two chemical engineering systems. The results of the GP-based identification when compared with those obtained using the transfer function based identification clearly indicates the outperformance by the former method.**

Keywords—process identification; predictor variable; genetic programming; sensitivity analysis; dynamic model

I. INTRODUCTION

Availability of an accurate, parsimonious and robust dynamical process model is essential in various tasks such as model based process control, monitoring and optimization. The field of building mathematical models of dynamical processes from the measured input-output data is known as "process/system identification." It can be viewed as the interface between the real world of applications and the mathematical world of control theory and model abstractions [1]. There are two principal ways for conducting process identification namely, *phenomenological* (first-principles) and *black-box*. In the first approach, the physico-chemical phenomena underlying a chemical process is rigorously described in terms of the mass, energy and momentum balance equations. This type modeling requires complete details of the governing phenomena, such as, kinetic rate constants, heat and mass transfer coefficients, and other thermodynamic information, which in most cases of practical interest are unavailable.

*Corresponding author email:vashinde@bvucoep.edu.in, vv.patil@ncl.res.in; BVDUCOE & CSIR-NCL, Pune.

Also, chemical processes very often exhibit complex nonlinear behavior, which makes the development of phenomenological models a tedious, costly and possibly even an impossible task to be completed in a reasonable time span. In such cases, the other approach i.e., black-box modeling is resorted to for process identification.

A black-box model representing the dynamics of a single input-single output (SISO) nonlinear process can be described using discrete time-variant inputs and outputs as given below:

$$y_{t+1} = f\left(y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-m+1}; u_t, u_{t-1}, u_{t-2}, \ldots, u_{t-n+1}\right) \quad (1)$$

where y_{t+1} refers to the *one-time-step-ahead* value of the output y, subscript t refers to the sampling instant, u is the manipulated variable (input), f denotes the functional relationship between y_{t+1} and the current and past (time delayed/lagged) values of the inputs and outputs, and m and n, respectively refer to the number of lags in the process output and input. In the above equation, the current and time-delayed inputs and outputs signify the predictor variables for the one-step-ahead-prediction of the output, i.e. y_{t+1}.

The principal advantage of the black-box modeling is that a model can be constructed solely from the measured process data without needing the details of the governing physico-chemical phenomena. In the conventional black-box modeling, the model structure is specified a-priori and the parameters associated with this model are estimated using an appropriate linear/nonlinear strategy. Since several efficient linear/nonlinear parameter estimation methods are available, the real difficult part in the black-box modeling is the specification of the model structure. For linear systems model specification is easy; however for a nonlinear system it poses significant difficulties since it involves selecting an appropriate model structure from the numerous competing ones.

The complexities involved in the conventional black box approaches to system identification necessitated exploration of alternative modeling strategies that do not require a-priori specification of the model structure. This requirement is fulfilled, for example, by a number of computational

978-1-4673-7992-2/16/$31.00 ©2016 IEEE