Computers and Electrical Engineering 72 (2018) 237–247

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Utilizing big data for batch process modeling and control

Abhinav Garg, Prashant Mhaskar*

Department of Chemical Engineering, McMaster University, Hamilton, Canada

A R T I C L E I N F O

Article history:
Received 30 March 2018
Revised 7 August 2018
Accepted 15 September 2018

Keywords:
Batch process
Subspace identification
Model predictive control
Big-data
Data driven model predictive control

A B S T R A C T

This manuscript illustrates the use of big data for modeling and control of batch processes. A modeling and control framework is presented that utilizes data variety (temperature or concentration measurements along with size distribution) to achieve newer control objectives. For an illustrative crystallization process, an approach is proposed consisting of a subspace state-space model augmented with a linear quality model, able to model and predict, and therefore control the particle size distribution (PSD). The identified model is deployed in a linear model predictive control (MPC) with explicit model validity constraints. The paper presents two formulations: (a) one that minimizes the volume of fines in the product by leveraging the variety of measurements and (b) the other that directly controls the shape of the particle size distribution in the product. The former case is compared to traditional control practice while the latter's superior ability to achieve desired PSD shape is demonstrated.

1. Introduction

In this age of digitalization, massive amount of data are generated by various sensing devices and is typically referred to as big data. It is generally characterized by 3 Vs namely, the requirements to handle large volumes, a variety of data, and data with high velocity (see, e.g., [1]). In some instances, additional parameters - establishing veracity and value, are included to characterize the challenges associated with big data. Several of these challenges (such as those related to handling large databases) are best addressed within the fields of computer science and business analytics applications. An aspect of big data that falls within the realm of process control problems is that of handling data variety and veracity.

In the context of process control, data variety would refer to availability of different types of process measurements such as temperatures and pressures data along with infrared spectroscopy, acoustic, or video/image data. The key distinction in these data types is in the structure of data. For instance, in a crystallization process, temperature and concentration measurements are collected with respect to time and is a two-dimensional dataset, however, data collected from spectroscopy (particle size distribution) is a three dimensional dataset with measurements varying both with respect to time and particle size. The challenge then lies in adapting the data-driven approaches to simultaneously handle such variety in modeling and eventually use it for control. In case the modeling approach is not able to simultaneously handle two dimensional and three dimensional measurements, alternative would be to reduce the particulate size distribution measurements (or any higher dimensional data) to two dimensional measurements [2]. Similarly, the data veracity refers to uncertainty in data which

* Corresponding author.
E-mail address: mhaskar@mcmaster.ca (P. Mhaskar).

https://doi.org/10.1016/j.compeleceng.2018.09.017

may stem from missing / corrupt measurements in data historian, non-uniform sampled measurements etc. The scope of the present work is restricted to handling data variety with application to batch crystallization process, and we make contributions both in the areas of modeling and control.

Model development generally takes one of two paths- so called first principles modeling, where the structure of the model is given by governing equations (appropriate balances, laws of thermodynamics), or data driven model, where the structure of the model is chosen based on simplicity of model development (albeit cognizant of the ability to predict process dynamics). Both models, however have parameters that need to be identified/determined, and this can only be done via measurements. In this sense, the availability of more data (or variety of data) generally increases the ability to capture the process dynamics better and thus presents an opportunity for advanced control implementation.

The presence of variety of data (henceforth referred to as data variety), however, poses a challenge with first principles model based control. In this approach, the model complexity increases in order to capture data variety. For instance in case of a particulate process, while the temperature and concentration measurements can be described using ordinary differential equations, prediction of particle size distribution requires population balance equations. Although results exists in the literature where first principles based approach have been successfully used for control (for instance see, [3]), use of alternative, data driven approaches (also referred to as data analytics) remains attractive due to the abundance of data and the associated simplicity of model development [4,5]. A key advantage over first principles based approaches is the resultant feasibility of online optimization based control approaches due to the use of linear models (thus often resulting in convex optimization problems), while still being able to adequately capture the complex dynamics. For instance, mid-course correction based control strategy for the control particle size distribution (PSD) in an emulsion semi-batch polymerization was proposed in [6] to drive it to a desired end-point. The approach utilized the available online and offline measurements for prediction of the final PSD using partial least squares (PLS) models, and then compute the required mid-course corrections, with the ‘timing’ of this mid-course correction based on heuristics. In [5], another PLS modeling approach was used for predicting the final PSD. The number of states of a distributed parameter process obtained after applying the discretization based solution strategies presented earlier for the solution of PBEs are very high, often resulting in ill-conditioned and uncontrollable systems. In order to overcome this, a principle component analysis (PCA) based model order reduction method was proposed in [7]. In this approach, a linear time varying (LTV) model was obtained by transforming the original model into a reduced order model which was subsequently used in the linear MPC framework to obtain the product with a desired PSD.

Over the past several decades, data analytics in process systems engineering has gained prominence. Largely, methods have been developed and applications demonstrated for traditional data (for instance, see [8]) with limited results illustrating robust archiving (volume) aspect of the big data in batch processes [9]. Some applications of neural network or machine learning based modeling and control of batch processes (not necessarily using data variety) have been demonstrated in the literature[10,11], however, largely machine learning based approaches have emerged for classification kind of problems such as in the area of fault detection [12–14].

Among other data-driven modeling approaches, latent variable technique is one method that has been used widely for modeling and control of batch quality. The simplest implementation of these methods uses batch-wise unfolding strategy with each row representing the batch duration. However, unequal batch duration is problematic with this unfolding strategy and inevitably requires batch alignment [9,15]. Therefore, as the mapping between future variable trajectories and real time domain are unclear, these approaches are more suitable for batch monitoring than control. A multi-model approach was proposed in [16] that did not require batch length alignment and utilized clustering algorithms to build the multiple models. Recently in [17], a batch quality modeling and control approach based on subspace identification was proposed that too doesn’t require batches to be aligned for building the model (due to ‘state’ dependence instead of ‘time’). While these recent results are based on building models using traditional data, they presents an opportunity to be adapted to handle the big data modeling problem.

Motivated by these considerations, this paper presents the framework to utilize big data aspects for modeling and control of batch particulate process using a dynamic subspace identification based approach coupled with a static product quality model. The key contributions of the present work are (a) a dynamic and quality model to handle both two dimensional and three dimensional measurements i.e. the data variety and (b) a novel MPC design with explicit model validity constraints. The rest of the paper is organized as follows: Section 2 presents first-principles based model and dynamics of seeded batch crystallizer. Section 3 outlines the proposed approach for handling data variety in the batch subspace identification framework for modeling and control of batch crystallizer. The section discusses two formulations: (a) minimizing the volume of fines in the final product by leveraging the variety of measurements and (b) control of shape of the particle size distribution in the product. The former case is compared to traditional control practice while the latter’s superior ability to achieve desired PSD shape is demonstrated. Finally, concluding remarks are made in Section 4.

2. Preliminaries

This section reviews a first principles model of a seeded batch crystallizer that will be used for validating the proposed big data based modeling and control approach.