Chemical Engineering Science 140 (2016) 44–61

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Comparative study of surrogate approaches while optimizing computationally expensive reaction networks

Srinivas Soumitri Miriyala, Prateek Mittal, Saptarshi Majumdar, Kishalay Mitra *

Department of Chemical Engineering, Indian Institute of Technology Hyderabad, Ordnance Factory Estate, Yeddumailaram 502205, India

HIGHLIGHTS

- Intelligently designed Sobol-ANN: A step towards online optimization of PVAc-LCB.
- Configuration of parsimonious Sobol-ANN obtained using novel MOOP formulation.
- Novel algorithm eliminates heuristic based bias towards monolayer perceptron NN.
- Parameter free ANN designing algorithm: requires only the model data as input.
- Sobol-ANN surrogate reduces 90% function evaluations; 1.5 times faster than KI.

ARTICLE INFO

Article history:
Received 8 December 2014
Received in revised form
31 March 2015
Accepted 23 September 2015
Available online 22 October 2015

Keywords:
PVAc-LCB reaction network
Genetic algorithms based online process optimization
Surrogate
Neural networks
Sobol
Kriging

ABSTRACT

Process modeling and optimization of polymerization processes with long chain branching is currently an area of extensive research owing to the advantages and growing popularity of branched polymers. The highly complex nature of these reaction networks results in a large set of stiff ordinary differential equations to model them mathematically with adequate precision and accuracy. In such a scenario, where execution time of the model is expensive, the idea of going for online optimization and control of such processes seems to be a near impossible task. Catering these problems in the ongoing research, optimization using surrogate model obtained from a novel algorithm is proposed in this work as a solution. A Sobol set assisted artificial neural network replaces the computationally expensive kinetic model of long chain branched poly vinyl acetate as the fast and efficient surrogate model. The proposed multi-objective methodology allows the computationally expensive first principle model to determine the configuration of the neural network, which can emulate it with maximum accuracy along with sample size required. The algorithm introduces a logical way of designing ANN architectures where the outperformance of multiple layer networks justifies the elimination of heuristics approach to consider only single layer. The results of the proposed algorithm are compared with the results obtained using Kriging interpolator based another surrogate approach, for testing, validation and scope of improvement. The use of fast and efficient Sobol assisted ANN surrogate model makes the optimization process ~10 times more efficient as compared to the case of optimization with computationally expensive kinetic model. The proposed ANN based surrogate is nearly 1.5 times as efficient as Kriging model in terms of number of expensive function evaluations.

1. Introduction

An extremely beneficial and pertinent class of polymers, namely the branched polymers, have marked a new era in the domain of polymer science and technology in the last few decades. Branched polymers have gained immense importance as compared to their linear counter parts due to the presence of a network structure, which affects the rheological properties in the downstream processing (Yan et al., 1999). The industrial applicability of branched polymers is such that in many cases, modified catalysts such as metallocene and bi-functional catalysts are used specifically to ensure branching in the reactors which otherwise do polymerization in general (Kolodka et al., 2003; Nagasawa and Fujimoto, 1972; Keramopoulos, 2002). Continuous endeavor is there to build mathematical models for such reaction mechanisms so that the operation can be optimized and controlled to harness the maximum benefit out of the system (Krajnc et al., 2001; Gretton-Watson et al., 2006). However, optimization of such

* Corresponding author. Tel.: +91 4023017055.
E-mail address: kishalay@iith.ac.in (K. Mitra).

http://dx.doi.org/10.1016/j.ces.2015.09.030
0009-2509/

reaction network based mathematical models of branched polymers requires large simulation time due to the complexity involved in the models (Mogilicharla et al., 2014).

On application of a method such as population balance to estimate the molecular weight distribution (MWD) of the polymer, the mathematical models comprise a large number (often few thousands) of stiff ordinary differential equations (initial value problems – ODE-IVPs), which may require long execution time (say, a couple of days) to have numerical solutions (Thomas, 1998). The approach of the moment based modeling for large system of equations is represented in terms of leading moments of different states of the system (Teymour and Campbell, 1994) and generally it reduces the complexity involved. However, since the MWD of the branches (moments) is not considered separately, the reconstruction of the full MWD from the equations involving those leading moments is not ensured directly with moment based modeling. Pladis and Kiparissides (1998) have reported that this problem can be resolved by assuming an analytical distribution function for the full MWD and then utilizing the leading moments to estimate the parameters appearing in the analytical function. This method, probably the most promising and efficient one, still remains to be computationally exorbitant when applied to highly complex processes (Mogilicharla et al., 2015). Moreover, the optimization of such models for finding the optimum operating conditions calls for multi-objective optimization formulations owing to the conflicts involved among the objective functions (Mogilicharla et al., 2014). The ability of evolutionary optimization techniques to solve multi-objective optimization problems is quite unmatched as compared to their counter classical optimization methods (Nain and Deb, 2002). However, in order to generate a wide spread conflicting Pareto optimal (PO) solutions, these evolutionary optimization techniques require a large set of candidate solutions (population) (Deb, 2001) for which the computationally expensive model has to be solved repeatedly. The multiple runs of those models necessarily cannot be avoided because any decrease in the population number may not actually lead to a high quality wide spread PO solutions (Deb, 2002). In such a scenario, where the online optimization of the process remains a forlorn, surrogate models can emulate the computationally expensive models accurately using a small and limited set of sampling data originated from the time expensive physics based model to make the optimization runs faster as a result (Mitra and Majumdar, 2011).

The philosophy of surrogate models is to generate accurate functional relationship among inputs and outputs of a given process. The most widely used surrogate models are function approximation models based on response surface methodology (RSM), artificial neural networks (ANN) and Kriging interpolators (KI) (Jin, 2005). RSMs are statistical models, which try to regress lower order (commonly, second order) polynomial models followed by conducting a sequence of designed experiments to guide the optimization search in a direction of optimal response of the objective function (Myers et al., 1989). These are one of the extensively used surrogate models applied across various domains (Dominguez-Perles et al., 2014; Bezerra et al., 2008; Myers et al., 2009). Several instances of failure in capturing the local surface utilizing lower degree polynomials led the RSM research into dealing with higher degree polynomials. KI, which has proved its immense scope of applicability in the areas of system identification, parametric analysis and optimization (Guanyu Zhang et al., 2013), geosciences (Mehdi Badel, 2011), statistics, design and analysis of computer experiments (Jones, 2001), is yet another popular function approximation technique. It uses Gaussian distribution functions to fit the training data with a set of parameters which can be tuned based on the estimation of potential error in interpolation. The interpolator predicts the output using the weighted combinations of predictions from simple basis functions (Jones, 2001).

On the other hand, ANNs are mathematical models, which try to mimic the functioning of biological neural network of human brain. They are widely acknowledged for their immense applications in pattern recognition problems, image processing (Yegnanarayana, 1994) and many other chemical engineering applications (Betiku and Ezekiel Taiwo, 2015; Karimi and Ghaedi, 2014; Reza Soleimania and Shoushtarib, 2013; Esmaeili and Dashtbayazi, 2014; David, 2000). The number of nodes in a single layer and the number of layers in the network together constitute the architecture of the network. The process of designing the optimal configuration of the network architecture of ANN often involves a method of hit and trial. Even after fixing the architecture, determination of the number of sample points for training, always leads to an impasse (Dua, 2010). Nuchitprasittichai and Cremaschi (2012) developed an algorithm to determine the sample size of a given network architecture using the K-Fold cross validation technique. Their work has delved upon the idea that when a fixed network topology gets trained with different number of sample points, it results in altogether a new set of parameters and thereby a different model. They then applied the K-Fold cross validation technique to each model and evaluated the cross validation error in terms of mean of deviations from each fold. The sample size increment was obtained by starting with a smaller size and then gradually incrementing it using incremental Latin Hypercube Sampling (i-LHS). The increment in sample size was terminated when the change in K-Fold error with respect to the change in sample size attained an acceptable tolerance value. However, the i-LHS way of incrementing the data set demands addition of new sample points and deletion of some already existing sample points to ensure that the newly emerged sample set conforms the latin hypercube structure at every step of the algorithm. This work, therefore, provides further scope of improvement in terms of (a) reducing the computational burden involved in determining the sample size and (b) providing balance between prediction accuracy and network complexity while determining the sample size. Many other researchers tried to train the weights by using the genetic algorithm to ensure an optimum hybrid ANN while Dua (2010) used the MINLP approach to obtain the optimal configuration of the neural network. On the other hand, Giri et al. (2013) have reported a procedure where a multi-objective optimization problem (simultaneous minimization of network complexity and maximization of accuracy) has been formulated to come up with an ANN topology in search of parsimonious models. However, they have restricted their study to a single layered fixed network topology without considering the optimal data size required to train a model. Although it has been customary to start with a single layered topology due to the assumption that single layered networks with sufficient nodes can predict almost all the non-linearity present in a data set, one cannot fix the number of layers in the architecture based on heuristics (Roy et al., 2006; Haykin, 1994; Hagen Howard et al., 2002). This is due to the fact that a single layer in ANN represents geometrically a hyperplane which tries to classify the given data into two sets. If the given data is very randomly distributed in the n-dimensional space resulting in a linearly inseperable data, it definitely needs more than one n-dimensional hyperplane for classification. This certainly invigorates the need for exploring multi-layered neural network topologies for obtaining better accuracy. However, almost no work has been reported in the literature that addresses these concerns of determining a simple ANN architecture along with the sample size required by it, which can predict results accurately with less computational burden.

In this article, catering the aforementioned needs, the authors propose a novel and computationally economical algorithm which