MERGING MACHINE LEARNING WITH MECHANISTIC MODELS VIA SEQUENTIAL AND INTEGRATED HYBRID PROCESS MODELING

William T. Bradley, Fani Boukouvala*
Georgia Institute of Technology
Atlanta, GA 30332

Abstract Overview

Hybrid or semi-parametric models are composed of physics-based equations and data-driven surrogate model components. Such models have shown great promise in the realm of process systems engineering due to their ability to accurately model a process where some of the system properties are known (i.e., continuity equations), but certain mechanisms are unknown or difficult to model (i.e., reaction or transport properties). However, despite the promise of hybrid modeling, the method has yet to find much application in practical industrial case studies. Lack of understanding of the hybrid modeling mechanisms and turn-key implementations of hybrid modeling techniques are the main reasons for lack of adaptation. An evaluation of different implementations of hybrid modeling methods via open source software would encourage the widespread use of hybrid modeling. Towards this end, we investigate two fundamentally different approaches to formulating hybrid models and evaluate their relative merit in terms of accuracy, efficiency, and generalizability to other systems. Using the sequential training procedure, the data-driven model is fitted without physical constraints, whereas when using the integrated approach, the data-driven model weights are optimized simultaneously with physical constraints obtained from the mechanistic model.

Keywords

Hybrid semi-parametric models, Neural Networks, NLP Optimization

Introduction

Increased computational power has brought machine learning to the fore of predictive modeling with its myriad empirical or data-driven techniques, achieving promising results in various fields of science and engineering. However, limitations inherent to machine learning (ML), such as poor extrapolation potential, raises interesting questions such as how to strike a balance between empirical and physics-based modeling for process systems engineering (PSE). Not surprisingly, the proper merging of ML data-driven tools with traditional, mechanism-driven approaches is not a new problem and has been the subject of study for many years (Thompson and Kramer, 1994; Psichogios and Ungar, 1992). Termed hybrid semi-parametric modeling, these approaches have the advantages over other model-building techniques in that a system-level understanding can be achieved while certain mechanisms (i.e., reaction rates, transport properties, the effect of processing conditions) remain unknown (von Stosch et al., 2014). This can be especially useful when developing the domain knowledge to accurately model the unknown mechanisms via classical mechanistic approaches is prohibitively expensive.

However, applying hybrid modeling to PSE is not without obstacles. The optimal structure of hybrid models

* To whom all correspondence should be addressed

can vary based on the end objective of the modeler and the development of appropriate surrogate models features its own domain of expertise with which the modeler often must become familiar. For example, the volume and accuracy of the available data may necessitate certain surrogate models over others. The burden imposed by these obstacles can be alleviated significantly by development of a framework wherein various hybrid models can be constructed and their performance validated. We present two fundamentally different approaches to hybrid modeling, namely the sequential approach and the integrated approach. Both approaches are developed using open-source software chosen for their high-throughput capabilities and thorough documentation. We compare the performance of the two approaches in terms of ease of implementation, accuracy and application flexibility. A hybrid modeling benchmark of fed-batch fermentation is used as a case study. In this case study, dynamic profiles of penicillin production are predicted via a combination of a mechanistic component (i.e., ODE continuity equations) and a data-driven component (i.e., rate equations). In the hybrid modeling literature, two structures have been proposed, namely the "serial" and "parallel" hybridization (von Stosch et al., 2014). In this work, we focus on the "serial" approach, which entails representing via a data-driven model only the unknown component (in this case the reaction rates). In the parallel architecture, the surrogate model corrects in parallel the predictions of the engineering model. However, in addition to requiring additional mechanistic knowledge, these "parallel" model configurations are known not to retain any extrapolation capabilities beyond a data-driven model (Van Can et al., 1996).

Methods

We make as our objective the modeling of a fed-batch penicillin production wherein the continuity relationships are known and observed, but the reaction rate mechanisms are unknown. A simulation of fed-batch penicillin fermentation is used to produce experimental data, consisting of reactor state variables biomass concentration (B), substrate concentration (S), penicillin concentration (P), and dilution rate (D). The model for reaction kinetics is borrowed from (Thompson and Kramer, 1994). Data is simulated for 9 fermentation batches for 200 hrs with samples taken at 3 hr intervals.

We then validate the performance of the resultant model by simulating 6 fermentation batches, the first using conditions from the training set, while the remaining 5 observe processing conditions not used to train the hybrid model (extrapolation). The performance of the hybrid model is quantified as the mean squared error of the model prediction and simulated 'experimental' data. Each approach is further tested in terms of its robustness to noisy data.

Sequential parameter estimation and fitting

In this approach, the system of mass balances, represented by ordinary differential equations, is first solved to yield the local rate values. The solution of the rate values can be found via any algebraic modeling language with numerical discretization capabilities. As a second step, a surrogate model is selected, in this case a neural network, which receives as inputs state variables B, S, P, D at time t and predicts the reaction rates during the period Δt. From the equations from Thompson and Kramer, it is clear that each rate is a function of B and S only. However, since such information is not always known a priori, we use all state variables as inputs. The weights of neural network models are found through the backpropagation algorithm L-BFGS and optimal neural network structure is identified through grid-search and cross-validation. Finally, the neural network is added to the ODE equations and tested on data simulated from the fed-batch model with different operating conditions.

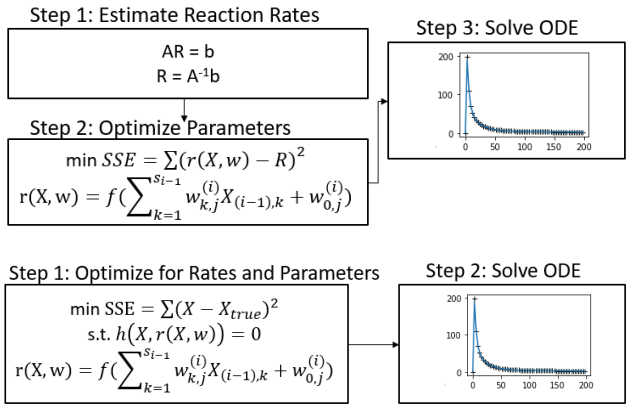

Figure 1. The sequential approach (top) first solves for the unknown rates, then fits the parameters of a data-driven model separately and sequentially. The integrated approach (bottom) solves for the rates and model parameters simultaneously.

Integrated parameter estimation and fitting

In the integrated approach, nonlinear local and global solvers are used to find the weight values of the neural network and the consequent values of the rates simultaneously. A single surrogate model is used, which in this case is a neural network with the same structure as that used in the sequential approach for easy comparison. A local nonlinear solver is used to find the parameters of the data-driven model—in this case, the neural network weights—that also satisfy the mass balance constraints. The performance of this model is evaluated by comparing its performance with fermentation batch data from different operating conditions than those used to fit the hybrid model. A comparison of the two approaches is depicted in Figure 1.

Results

Data used to test predictive capabilities of the hybrid models include operating conditions within the range of training data (interpolation) and outside the range (extrapolation). Due to space limitations, only the more challenging extrapolating case is shown here (Figure 2), since the predictions of the interpolating cases match the experimental data with high accuracy. The hybrid models are tested for extrapolation potential by predicting reactor performance when the initial biomass and feed concentrations are higher than those conditions used to fit the models. As observed in Figure 2, overall the sequential model's predictions are closer to that of the true data. However, in the more nonlinear regions, prediction deviate significantly, a sign of overfitting. Moreover, in the case of penicillin predictions, the sequential hybrid model predicts negative product concentration during the initial hours of fermentation, violating physical constraints. Meanwhile, the integrated approach yields a hybrid model that violates non-negativity to a lesser degree.

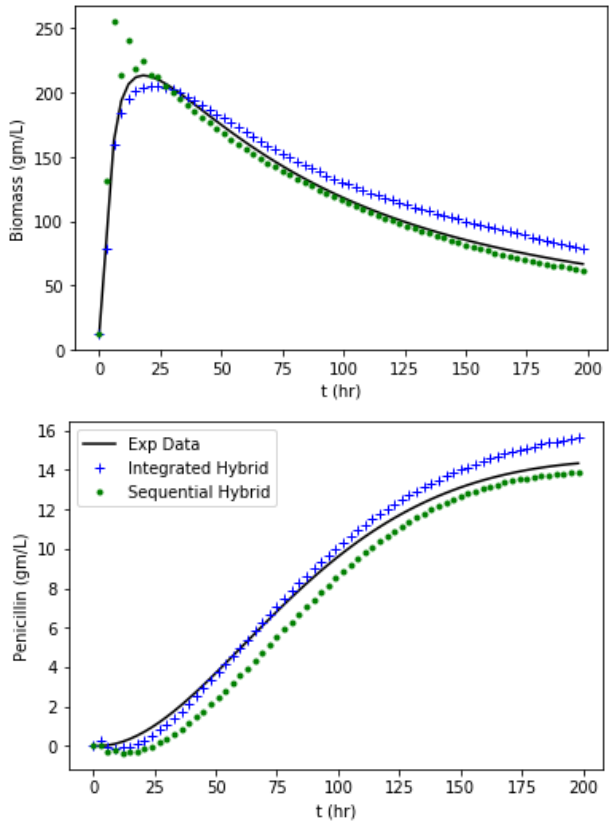

Figure 2. Biomass (top) and penicillin (bottom) predictions from two approaches to hybrid modeling.

Conclusions

The two approaches explored above represent different modeling paradigms that govern their applicability. In the sequential approach, modeling tools more familiar to the ML community are used to systematically vet various surrogate models before selecting an optimum via cross-validation. This enables the user to make a straightforward comparison of multiple surrogate model types and structures. Unfortunately, this method's use of backpropagation can guarantee only a locally optimal solution even for the 'best' model. In contrast, the integrated approach exploits nonlinear programming theory to solve for both the rates and the neural network parameters simultaneously, and, if a global solver is used, find a globally optimal solution. The potential guarantee of global optimality is unique to the integrated approach and needs investigation. However, it should be noted that computational requirements of this approach limit the exploration of possible neural model models to small structures for most applications of interest.

In future work, we will investigate the robustness of the two approaches and explore more rigorously the merits and limitations of the two approaches in terms of modeling effort, computation time and prediction accuracy.

Acknowledgments

The authors wish to thank Georgia-Tech start-up grant for financial support of this work.

References

Psichogios, Dimitris C., and Lyle H. Ungar. 1992. 'A hybrid neural network-first principles approach to process modeling', *Aiche Journal*, 38: 1499-511.

Thompson, Michael L., and Mark A. Kramer. 1994. 'Modeling chemical processes using prior knowledge and neural networks', *Aiche Journal*, 40: 1328-40.

Van Can, Henricus J. L., et al. 1996. 'Strategy for dynamic process modeling based on neural networks in macroscopic balances', *Aiche Journal*, 42: 3403-18.

von Stosch, Moritz, et al. 2014. 'Hybrid semi-parametric modeling in process systems engineering: Past, present and future', *Computers & Chemical Engineering*, 60: 86-101.