PHYSICALLY CONSISTENT DATA-DRIVEN SOFT-SENSOR DEVELOPMENT

Yu-Da Hsiao[1], Cheng-Hung Chou[1], Hai-bin Wu[2], Jia-lin Kang[3], David Shan Hill Wong[1], Yuan Yao[1], Yao-Chen Chuang[4], Shi-Shang Jang[1,4], John Di-Yi Ou[4]

[1]Department of Chemical Engineering, National Tsing Hua University, 30013, Hsinchu, Taiwan
[2]Graduate Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[3]Department of Chemical Engineering, Tamkang University, Taipei, Taiwan
[4]Center for Energy and Environment Research, National Tsing Hua University, 30013, Hsinchu, Taiwan

Abstract

Recent advances in artificial intelligence and data analytics have demonstrated that complex input-output relation can be learned by a deep neural network architecture when a large amount of training data is available. In development of data-driven soft-sensor for process control, there is always the concern whether sufficient physics were captured to allow for successful extrapolation and inference of other properties. To make up for the lack of important information, a dynamic model using past sensors readings and operating parameters was commonly used. However, simple input-output models in the form backpropagation neural work were shown to be unable to predict correct differential change of output (i.e. process gain), if the amount of data is not sufficient. To overcome this problem, one approach is to train a preliminary model using simulation data and use transfer learning to match plant data. It was found that gain consistency of the preliminary model can be preserved, while the prediction accuracy of the model can be substantially improved with only a limited amount of actual data. Alternatively, a semi-supervised training approach can be used. A nonlinear state-space model can be formulated in an observer-predictor form using a sequence-to-sequence recurrent network encoder. The observer can be trained using the large amount of unlabeled data provided by flow, pressure and temperature sensors. This observer-encoder outputs a limited number of hidden states using unlabeled dynamic data in the past operating window. These hidden states and the current values of manipulated variables were then used as input for a predictor which can be developed with a limited number of labeled data. The gain sign is mostly consistent, indicating that sufficient physics was captured by the model.

Keywords: Soft-sensor, gain consistency, transfer learning, sequence-to-sequence recurrent network.

Introduction

Data-driven soft-sensor is a common practice in the chemical industry to improve quality control and process control (Kadlec et al 2009). Chemical processes are usually nonlinear in nature, hence the use of neural networks for soft sensor development (Gonzaga et al 2009, Rani et al 2013). Neural network were known to have capabilities to model complex input-output relations (Hornik et al 1989, Lu et al 2017) with sufficient data. There is always however, concerns whether such models have capture enough physics so that they can be extrapolate and generalized with confidence if there is not enough data or some critical information is not available.

Motivating Example

Consider a distillation column in a local refinery. The distillate composition of the heavy key y_t was measured once a day. Data of 11 manipulated variables $\boldsymbol{u}_t$, 8 sensors variables $\boldsymbol{sv}_t$, temperatures, pressures, flows and levels were available every minutes over a three year periods. During this period only 900 samples of distillate composition were available. Moreover, the feed composition was not also available. The reflux rate was operated manually to control distillate composition. It is desirable to build a soft-sensor for predicting distillate composition so that the composition control can be closed automatically.

To overcome the lack of information of the distillate model, a nonlinear input-output model was developed using online sensor data averaged every 10 minutes and a window width of 1 hours. Hence a simple backpropagation network with 1-output and $11 \times 6 + 8 \times 5 = 106$ inputs was constructed.

$$y_{t+1} = f(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}, \boldsymbol{sv}_{t-1}, \cdots \boldsymbol{u}_{t-W}, \boldsymbol{sv}_{t-W}) \quad (1)$$

The out of the 818 data, 103 were used as the test data and the BPNs were trained with 715, 358, 179, 90 samples, 1/10 of these samples were used as validation data.

The RMSE of test data is a commonly used indicator of the prediction ability of a data-driven model. An alternative way to test whether our data-driven model is consistent with the physics of the process is to calculate the process gain for some important operating parameters. Without loss of generality, let $\Delta u_{1,t,}$ be the change of reflux rate. The effects of changing the reflux rate at any data points were calculated for all the data:

$$K_t = \left(\frac{\Delta y_{t+1}}{\Delta u_{1,t,}}\right)_{(u_{2,t}, \cdots u_{11,t}), \boldsymbol{sv}_t, \cdots \boldsymbol{u}_{t-W}, \boldsymbol{sv}_{t-W}} \quad (2)$$

The quality variable is the composition of the light key in the distillate. Hence the sign of the gain must be negative for all samples for the model to physically meaningful. The gain consistency can be defined as the fraction of samples points at which the gain estimated by the model has the correct sign. Figure 1 plotted the relation between RMSE of test data and gain consistency. It is obvious that the gain consistency became very poor when the number of training data became less. Furthermore, there is no correlation between the RMSE and gain consistency.

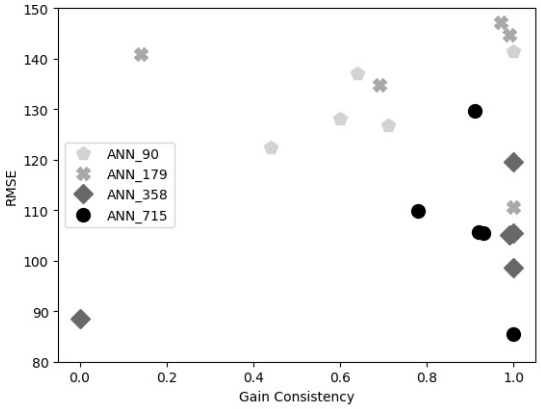


Figure 1: Relation between gain consistency and prediction accuracy of different BPN models

In some applications, the soft sensor model also serves as the input-output model for use in model predictive control. It is imperative that the gain sign is consistent with the actual gain such applications. Two approaches were investigated in the following.

Transfer Learning

For a simple process such as distillation, dynamic simulation data can be easily obtained using dynamic process simulators such as CHEMCAD. However, the actual dynamics may be different when the actual valve sizing, pipe sizes, tray internals, sump size and controller parameters are different. To obtain all these information may also be a tedious procedure. In this work, a dynamic process simulator was built by CHEMCAD. All equipment parameters such as column diameters, and valve size were determined using auto-sizing procedure. Controller parameters were obtained by auto-tuning procedure. Large amount of data were generated by random variations of inlet feed flow and compositions. Preliminary models (Table 1) with the same input-output structure as in equation (1) were trained using these simulation data. Due to the large number of training data available, we can see that the gains of these preliminary models were always gain consistent. However, these preliminary models perform poorly when used to predict actual plant data, as found in the last column of Table 1. This is understandable since all the equipment data and control parameter relevant to the dynamic behavior were not correct.

A transfer learning procedure known as fine tuning can be used to modify the model by retraining the weights of selected layers using available plant data (Yosinski et al 2014). In Table 2 show the prediction accuracy of the preliminary model can be substantially improved when it was fine-tuned by different numbers of training data, while retaining the gain consistency. However, the effects of fine tuning different weights are different, especially with a deep network. it should be noted that there is no general rule of thumb in transfer learning whether shallow weights or deep weighs should be fine-tuned (Yosinski et al 2014, Li et al 2016).

Table 1: Prediction accuracy of proxy models

Structure	#Param./ #data	RMSE Test Data	% Gain Correctness All data	RMSE Plant. Data
1 layer	1.20	157	100%	10562
2 layers	1.82	126	100%	4898
5 layers	3.04	111	100%	2957

Table 2: Prediction accuracy of proxy models

Number of Hidden Layers	Fine-tuned Weights	# of Training Data	#Param. / #data	% Gain Correct-ness All data	RMSE Plant. Data
1-layer	Hidden Layer 1 to Output	90	132	100%	109
		179	66	100%	112
		385	33	100%	101
		719	17	100%	95
2-layer	Hidden Layer 2 to Output	90	137	100%	154
		179	69	100%	90
		385	34	100%	82
		719	17	100%	92
5-layer	Hidden layer 5 to Output	90	137	73%	166
		179	69	96%	145
		385	34	91%	142
		719	17	99%	144
5-layer	Input to Hidden layer 1	90	131	100%	97
		179	66	100%	101
		385	33	100%	87
		719	17	100%	96
5-layer	Hidden Layer 1 and 5	90	268	100%	99
		179	135	100%	96
		385	67	100%	86
		719	34	100%	89

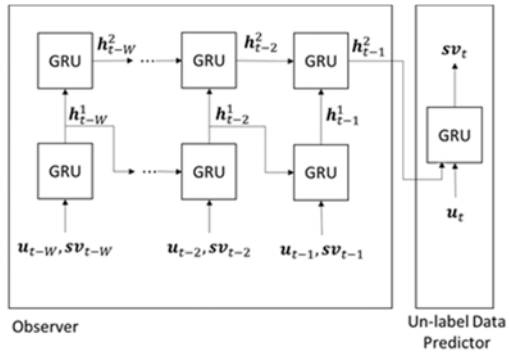
Sequence-to-Sequence Recurrent Network

A soft-sensor model can be cast into a sequence-to-sequence model, which is often used in speech recognition (Cho et al 2014; Sutskever et al 2014). The many-to-one observer model and the un-label data predictor Figure 2a can be trained using the large number of sensor variables available (80000 samples over a two-year period). After training the hidden state observer, the un-label predictor is replaced by a back propagation network (BPN), which uses the observed hidden states and the current manipulated variable as input, to predict the quality (Figure 2b).

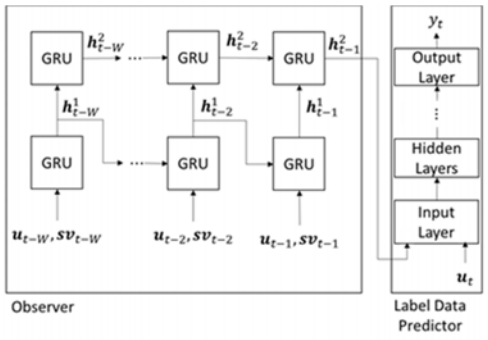
The relation between gain consistency and prediction accuracy of different label data predictor models obtained with various sizes of training data was shown (Figure 3). Regardless of the number of data used and structure of the neural network predictor, the gain consistency is always high.

The leverage of unlabelled data in machine learning is known as semi-supervised training method (Liu et al 2018). In model based soft-sensors, state variables of a physical model were determined using observers (Tham et al 1989). However, in our sequence-to-sequence model, no physical models were required. Both the observers and predictor

were data-driven. Unlabelled data were used to the train the observer in a supervised manner in which past history was used to predict future dynamics.



(a)The RNN observer and the unlabelled data predictor



(b) The ANN predictor

Figure 2: Structure of a sequence-to-to sequence recurrent network for soft sensor implementation

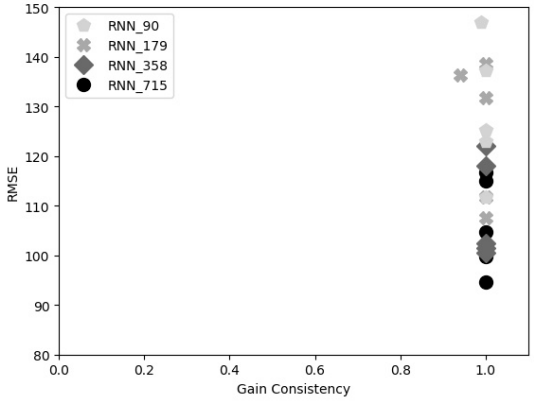


Figure 3: Relation between gain consistency and prediction accuracy of different label data predictor models obtained with various sizes of training data set

Conclusions

In this study, the effect of data availability in capturing physics of a process using a data-driven model were examined. Prediction errors of test data may not be a good indicator of whether the model is consistent with physics. The consistency of sign of the derivative of output to input is a more demanding indicator. Two approaches were presented to resolve the challenge of building a gain consistent model with limited data. If an approximate physical model is available, simulation data can be used to construct a base model, which can be transferred into a real model by fine tuning with a limited amount of data. Alternatively unlabeled data can be used to encode past history into a limited number hidden states in a sequence-to-sequence recurrent neural network, which can be then used to predict label data. In the application to a distillation example, we found that both approaches are able to improve the physical meaningfulness of the data-driven model.

Acknowledgement

The authors, except H.B. Wu, acknowledged the financial support of Ministry of Sciene and Technology of Taiwan through grant 107-2622-8-007-015-.

References

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width. In Advances in Neural Information Processing Systems (pp. 6231-6239).

Li, Y., Wang, N., Shi, J., Liu, J., & Hou, X. (2016). Revisiting batch normalization for practical domain adaptation. International Conference on Learning Representations (2016).

Liu, Y., Yang, C., Gao, Z., & Yao, Y. (2018). Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes. Chemometrics and Intelligent Laboratory Systems, 174, 15-21.

Kadlec, P., Gabrys, B., & Strandt, S. (2009). Data-driven soft sensors in the process industry. Computers & chemical engineering, 33(4), 795-814.

Gonzaga, J. C. B., Meleiro, L. A. C., Kiang, C., & Maciel Filho, R. (2009). ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. Computers & chemical engineering, 33(1), 43-49.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Advances in neural information processing systems (3104-3112).

Tham, M. T., Morris, A. J., & Montague, G. (1989). Soft-sensing: a solution to the problem of measurement delays. Chemical Engineering Research and Design, 67(6), 547-554.

Rani, A., Singh, V., & Gupta, J. R. P. (2013). Development of soft sensor for neural network based control of distillation column. ISA transactions, 52(3), 438-449.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, 27, 3320-3328.