Neural network and Sparse Identification of Nonlinear Dynamics Integrated Algorithm for Digital Twin Identification

Jingyi Wang *, Jesús Moreira **, Yankai Cao *, R. Bhushan Gopaluni *

July 9th, 2023

* Department of Chemical and Biological Engineering

University of British Columbia

** Digital, Innovation and Lean Capability Development Team

Imperial Oil Company

IFAC 2023

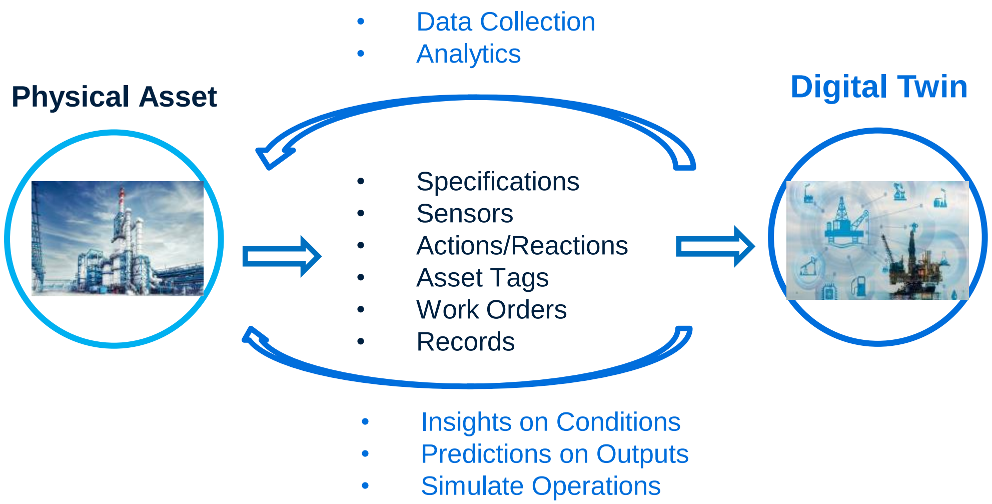
Outline

- Background introduction
- Neural network and sparse identification of nonlinear dynamics integrated algorithm
- Case study
- Conclusions

Background Introduction

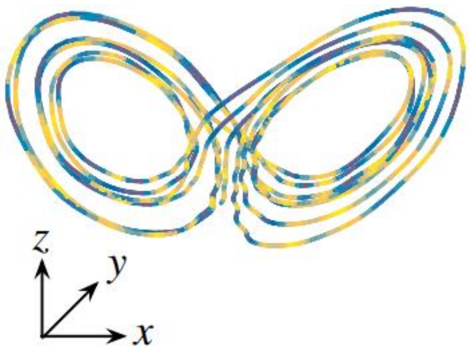
Digital twin

A digital twin is a computer-based mathematical representation that simulates the behavior of a given process. Digital twins are used to interact with and simulate real-world processes [1].

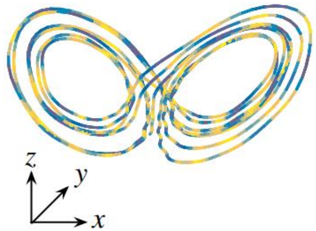


Sparse identification of nonlinear dynamics (SINDy)

The SINDy uses a three-step sparse regression framework to automatically discover the underlying governing equation of dynamic systems from the process data [2].



Sparse identification of nonlinear dynamics (SINDy)



Unknown dynamics

Data collection

$$\boldsymbol{\Theta}(\mathbf{X})$$

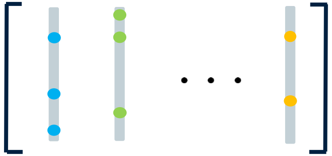
$$\begin{bmatrix} 1 & x_1 & \dots & x_1^2 & \dots & x_n^5 \end{bmatrix}$$

A library of candidate model terms

Solve sparse regression problem

$$\mathbf{Y} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi}$$

$$\begin{bmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{bmatrix}$$



$$\boldsymbol{\Xi}$$

Identified system models

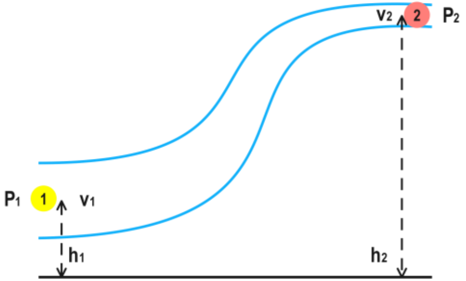
Feature generation

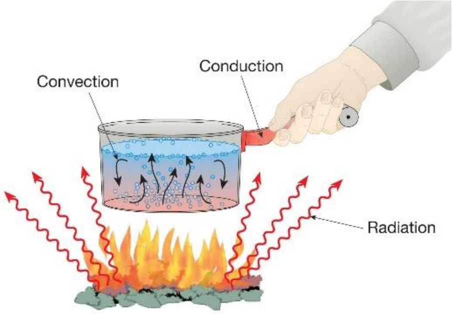
Data-driven feature generation:

$$\mathbf{\Theta}(\mathbf{X}) = [\mathbf{1} \;\; \mathbf{X} \;\; \mathbf{X}^{\mathrm{PO}_2} \ldots \sin(\alpha\mathbf{X}) \ldots \tanh(\beta\mathbf{X})]$$

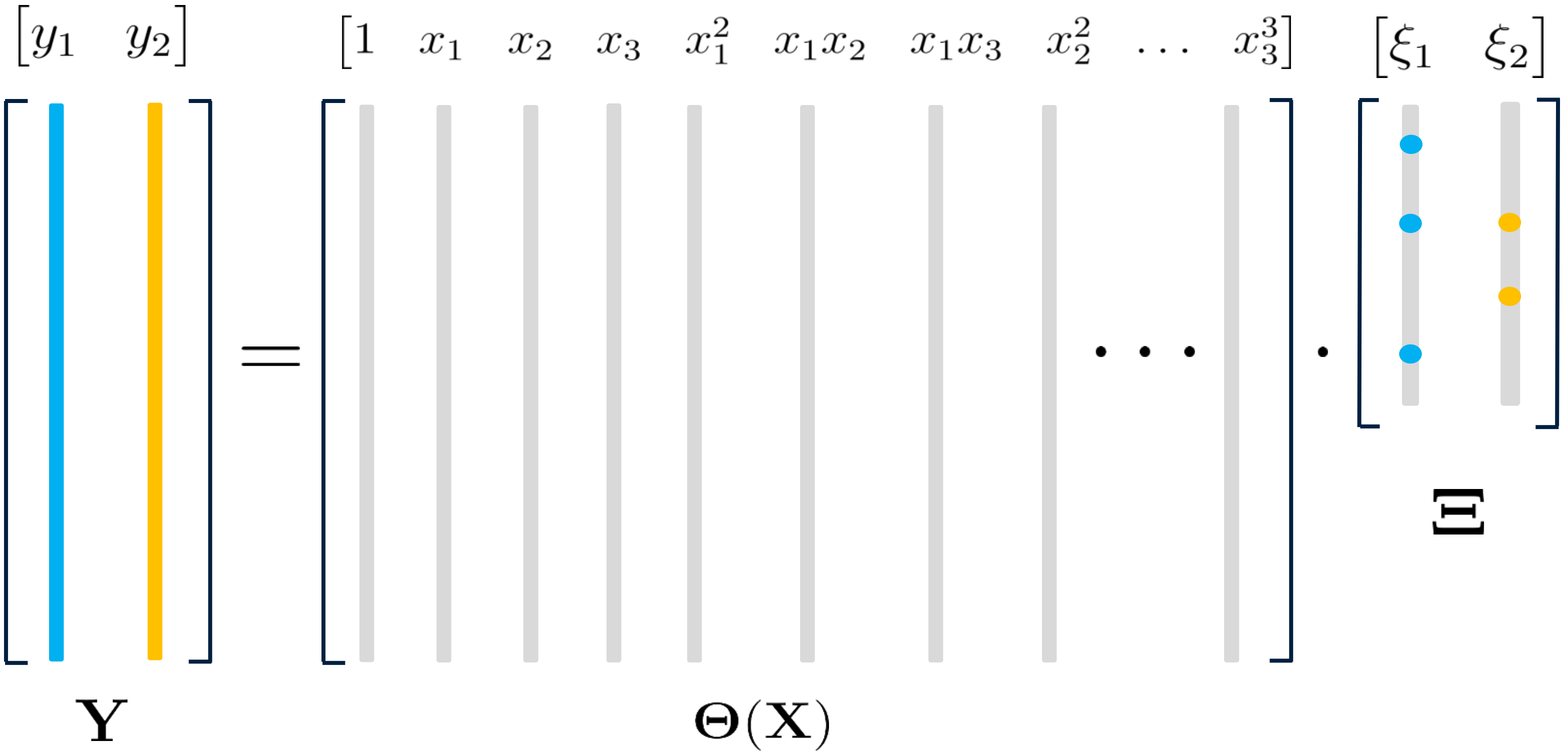
Hybrid feature generation:

$$\mathbf{\Theta}(\mathbf{X}) = [\mathbf{1} \;\; \mathbf{X} \;\; \mathbf{X}^{\mathrm{PO}_2} \;\ldots\; \rho g(h_2 - h_1) \;\; \mathrm{UA}\Delta T \;\ldots]$$





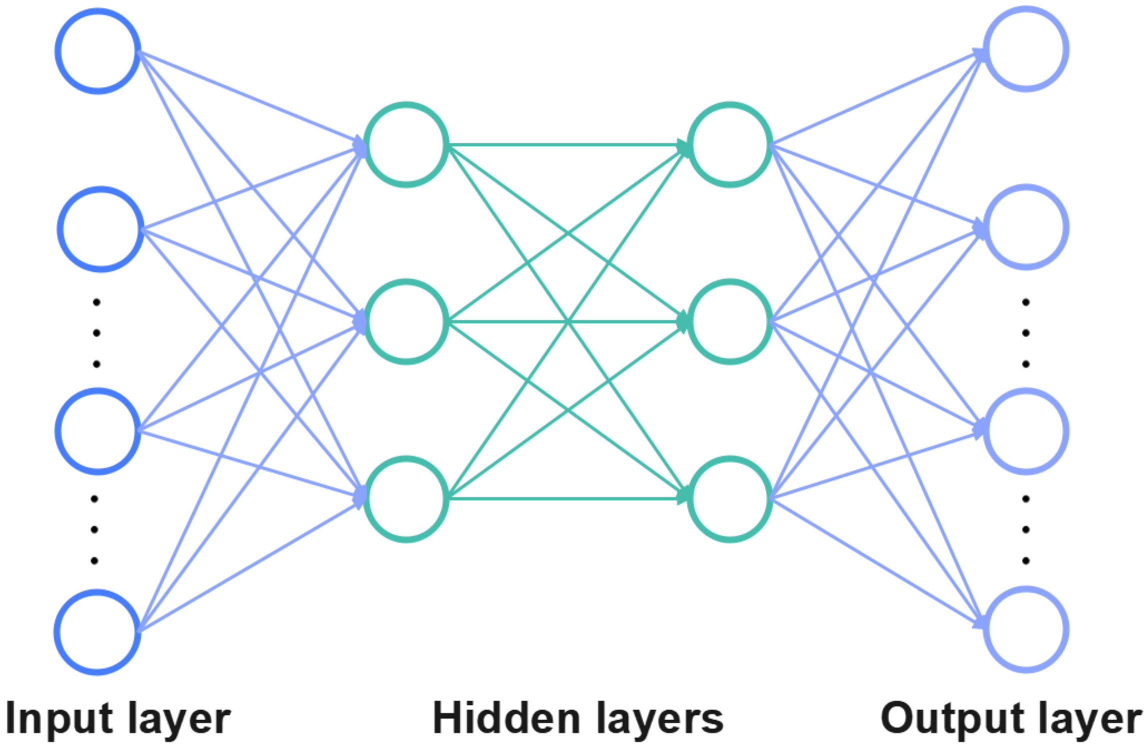
Feature selection



Research Motivations

1. SINDy has been applied to successfully identify governing equations in various fields, including to fluid dynamics [1], biology [3], chemical process [4], etc.
2. SINDy's feature engineering procedure allows users to combine first-principles information and data-driven techniques seamlessly.
3. SINDy is limited to identify linear-in-parameter relationships.
4. Artificial neural networks are more applicable to capture complex nonlinear relationships, such as rational relationships and implicit relationships.

Artificial Neural Network

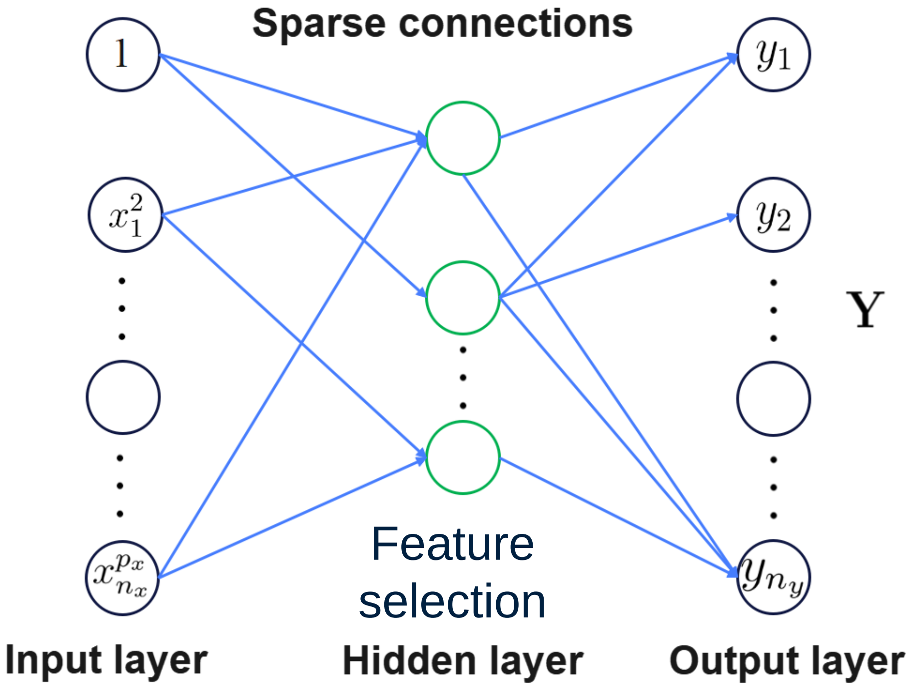


Neural network and SINDy integrated algorithm for digital twin identification

Neural network and SINDy integrated algorithm for digital twin identification

Feature generation

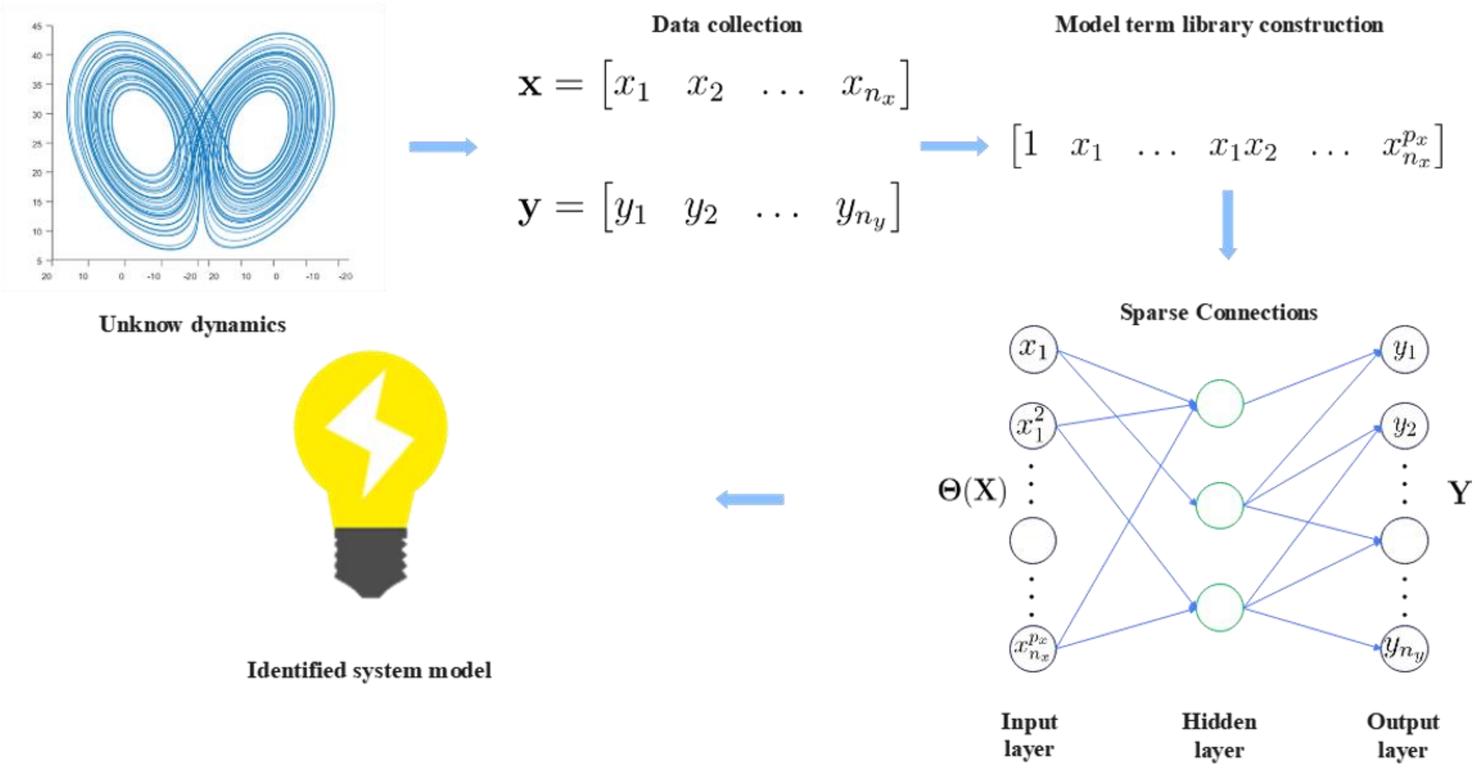
$\Theta(\mathbf{X})$



$\mathbf{Y}$

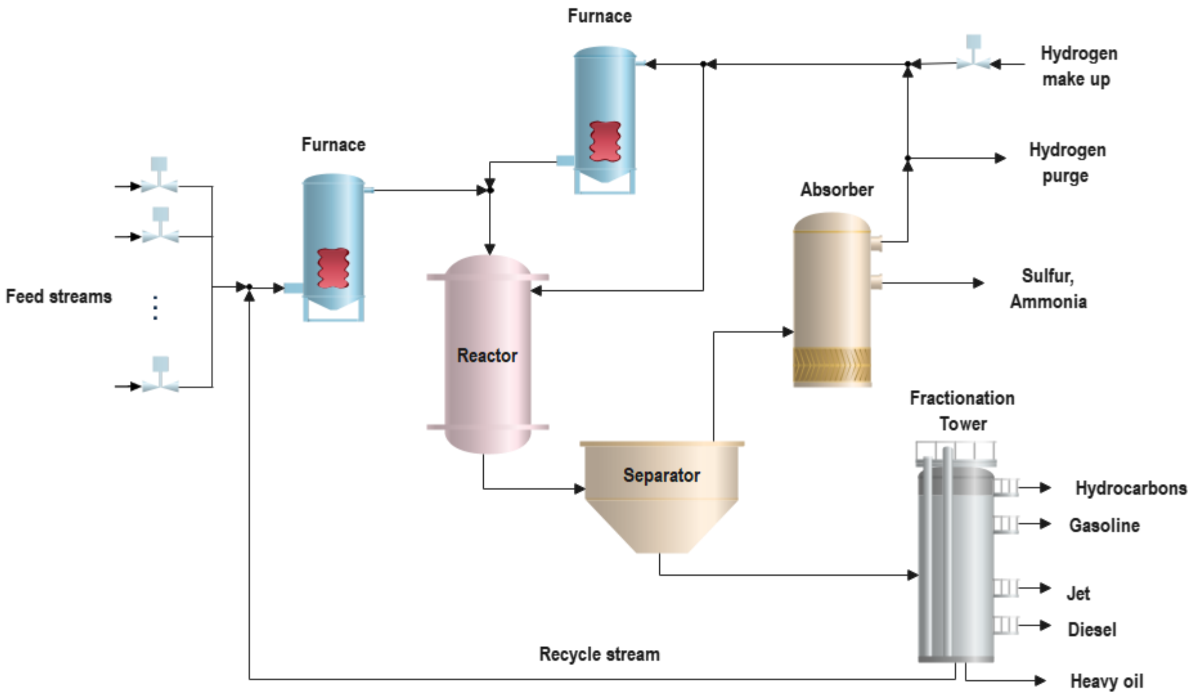
Digital twin identification

Neural network and SINDy integrated algorithm for digital twin identification



Case study

Diesel hydrotreating (DHT) unit example



Diesel hydrotreating (DHT) unit example

Data availability

	Real operational data	First-principles data
Sample number	8161	142
Number of input variables	37	13

Diesel hydrotreating (DHT) unit example

Feature generation

$$\mathbf{FP}(\mathbf{X}^{\mathrm{PO}_2})$$

$$\frac{1}{\rho}\ \mathrm{e}^{-\frac{\mathrm{E_a}}{\mathrm{RT}}}$$

$$\mathbf{\Theta}(\mathbf{X}) = \left[\mathbf{1}\ \ \mathbf{X}\ \ \tanh(0.8\mathbf{X})\ \ \mathrm{e}^{(\mathbf{X})}\ \ \mathbf{FP}(\mathbf{X}^{\mathrm{PO}_2})\ \ \frac{1}{\rho}\ \mathrm{e}^{-\frac{\mathrm{E_a}}{\mathrm{RT}}}\right]$$

Diesel hydrotreating (DHT) unit example

Performance comparison among the SINDy, the conventional single-layer neural network, and the proposed algorithm in terms of MSE

	Output yields (BPH)		
Methods	Gasoline	Diesel	Jet
SINDy	0.096	0.237	0.0839
Conventional single-layer neural network	0.068	0.2410	0.110
Proposed algorithm	**0.058**	**0.202**	**0.068**

Conclusions

- A neural network and SINDy integrated algorithm is proposed to automatically identify the complicated nonlinear digital twin model combining both first-principles knowledge and data-driven techniques.

- Compare to SINDy, the proposed method is more applicable to identify large-scale complex nonlinear relationships.

- Compare to traditional neural networks, the proposed approach combines both first-principles knowledge and data-driven techniques to improve the prediction accuracy and prevent overfitting.

Acknowledge

- Natural Sciences and Engineering Research Council of Canada (NSERC)
- Imperial Oil
- University of British Columbia

References

- [1] M.Day, "Discussing digital twins,"[Online]. Available: https://aecmag.com/features/discussing-digital-twins/, 2020.
- [2] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proceedings of the National Academy of Sciences, vol. 113,no. 15, pp. 3932–3937, 2016.
- [3] Mangan N M et al., 2016. Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics. IEEE Trans. Mol. Biol. Multi-Scale Commun. 2, (1):52–63.
- [4] Bhadriraju B, Narasingam A, and Kwon J S.-I, 2019. Machine learning-based adaptive model identification of systems: Application to a chemical process. Chem. Eng. Res. Des. 152,372–383.

Thank you and Questions?