Integrating Scientific Theory with Machine Learning

Miodrag Bolic

University of Ottawa
School of Electrical Engineering and Computer Science,

December 15, 2020

Outline

- Introduction
- Merging mechanistic and machine learning models
  - How is the knowledge represented?
  - Where is the knowledge integrated in the machine learning pipeline?
  - How is the knowledge integrated [1]?
- Integrating state-space and deep learning models
- Some works on hybrid models in bioreactors

Our interests

- Please note that my group is not in favor of black-box learning
- We are interested in:
  - generative and probabilistic models,
  - integrating physical models into ML,
  - working with time series (not i.i.d) data,
  - quantifying uncertainties in predictions and classifications,
  - providing some explainability.

Current approaches

Machine learning

- Not enough data to train sufficiently generalized models.
- Purely data-driven model might not meet constraints such as dictated by natural laws, or given through regulatory guidelines.
- Machine learning models becoming increasingly complex -> need for models to be interpretable and explainable
- Require homogeneous labeled training data

Mechanistic models

- "A picture is worth a thousand words" -> "A model is worth a thousand datasets."
- Normally, cannot capture complex dynamics in the system
- Often, they are simplified to allow for handling complexity

Hybrid models

- Best of both worlds
- Can be complex and difficult to train

Presenting knowledge

- *General knowledge*: knowledge independent of the task and data domain.
- *Domain knowledge*: knowledge in any field such as physics, chemistry, engineering, and linguistics with domain-specific applications.

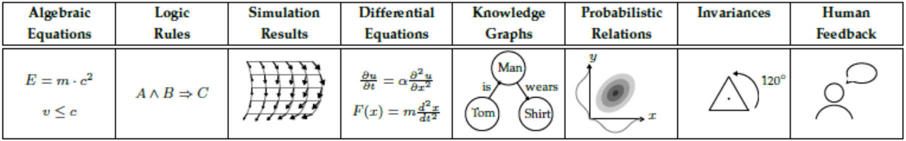


Figure: Domain knowledge representation [1] [1]

[1]Note that all figures are copied and the sources are referenced!

Information Flow of Informed ML

Informed machine learning describes learning from a hybrid information source that consists of data and prior knowledge. The prior knowledge is pre-existent and separated from the data and is explicitly integrated into the machine learning pipeline [1].

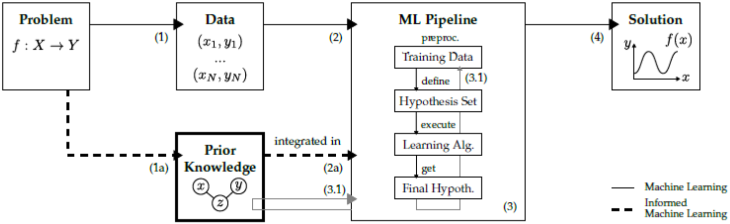
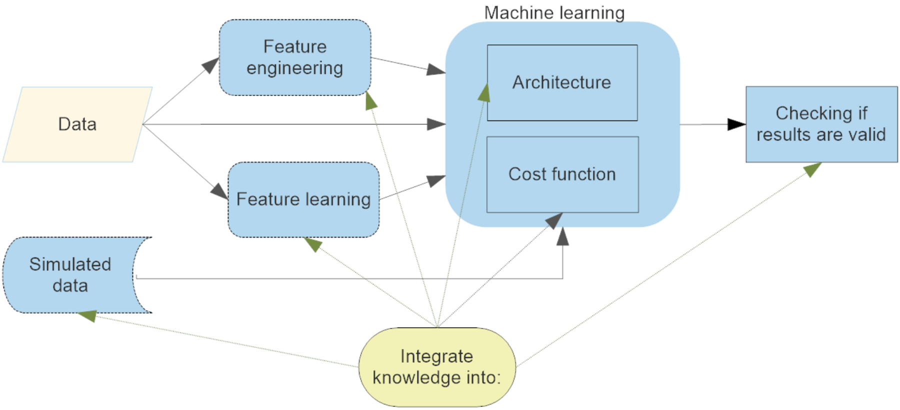


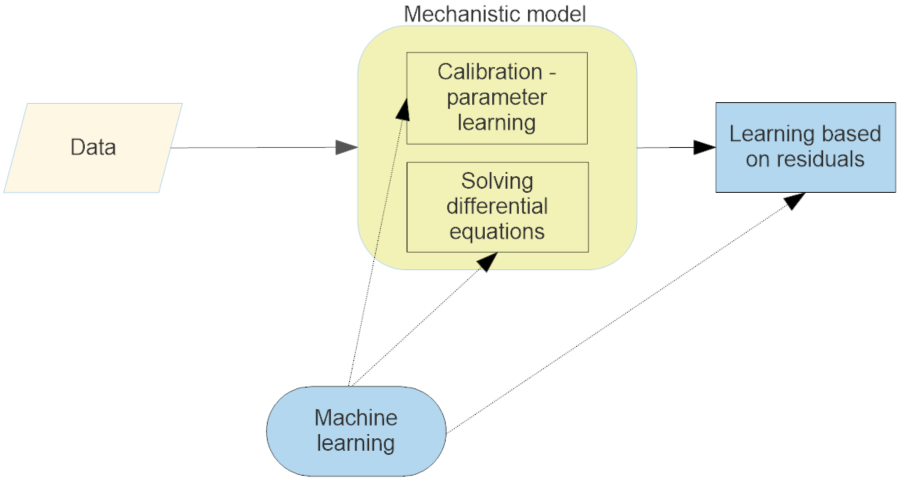
Figure: Machine learning flow [1]

Next, we will explore how one can integrate mechanistic models with machine learning.

Knowledge informed ML



Mechanistic models augmented with ML



Data and Training

Training data

- Features
  - feature engineering
- Data augmentation
  - image transforms
  - simulations: generate a large amount of data from mechanistic models for training.

Feature learning

- Unsupervised learning - knowledge can still be incorporated
- Variational autoencoders
  - VAEs jointly learn an inference model and a generative model, allowing them to infer latent variables from observed data.

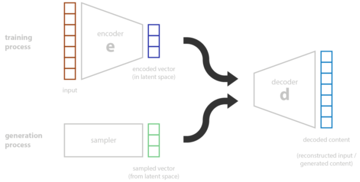


Figure: Understanding Variational Autoencoders (VAEs)

Residual modeling

ML models the residuals of the domain knowledge model and tries to reduce the error between the mechanistic model output and the ground truth.

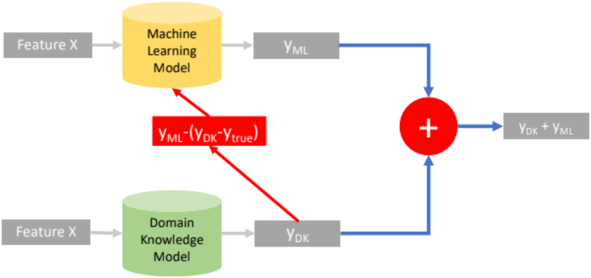


Figure: Residual modeling [2]

- Y_{ML}: machine learning predicted label
- Y_{DK}: domain knowledge predicted label
- Y_{true}: ground truth label

Learning algorithm

- *Mass* = *Density* · *Volume*: ML does not know that this is not supposed to be violated
- Domain knowledge is into a loss function and performs regularization

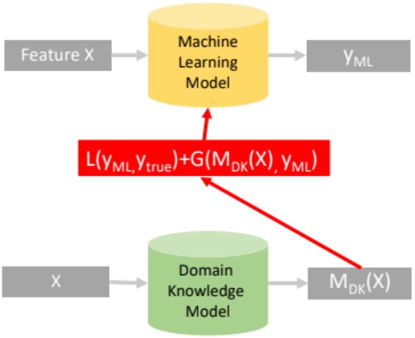


Figure: Knowledge in the loss function [2]

- function $G()$ is regularization term: a measure of consistency between domain knowledge and predicted label
- function $M_{DK}()$: a domain knowledge transformation of feature X

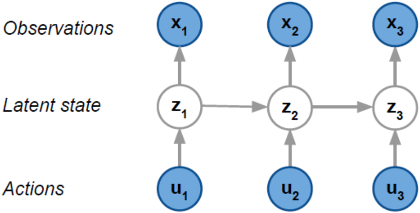
Architectures (incomplete)

- Model structure incorporates the mechanistic model
- We will introduce state-space models first and show how they can be integrated with RNNs
- Integration is done with variational autoencoders

State-space models

State-space models:

- are numerically efficient to solve,
- can describe differential equations,
- allow for a more geometric understanding of dynamic systems, and
- form the basis for much of modern control theory



- Transition equation:

$$\mathbf{z}_t = \mathbf{A}_t \mathbf{z}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \epsilon_t$$

- Emission equation:

$$\mathbf{x}_t = \mathbf{C}_t \mathbf{z}_t + \boldsymbol{\delta}_t$$

Figure: Linear state-space model [3]

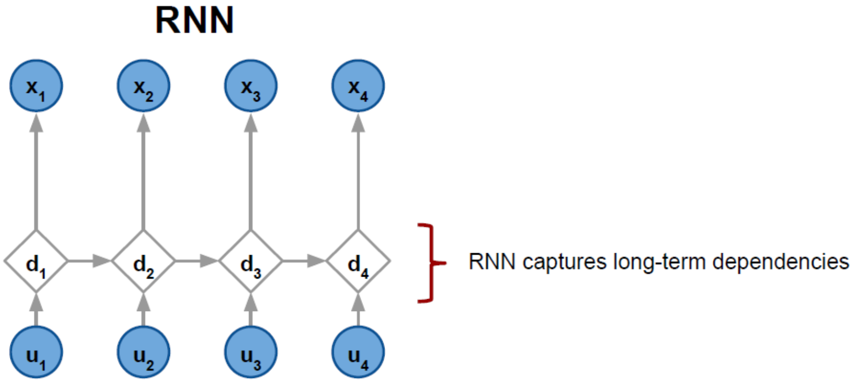
Filtering in state-space models

Kalman filter:

- Kalman filter is optimal for linear Gaussian problems.
- Generalizes many common time-series models
- Strong modelling assumptions:
  - Linear transitions and emissions
  - Gaussian transitions and measurement noise

Non-linear filters

- Extended Kalman filters (non-linear observation equation, Gaussian noise)
- Particle filters (non-linear, non Gaussian)
- Problems
  - Transition model still have difficulties handling complex non-linear dynamics
  - Does not capture long-term dependencies in data (Markov models)



RNN
x1
x2
x3
x4
d1
d2
d3
d4
u1
u2
u3
u4
RNN captures long-term dependencies

Figure: RNN [3]

Stochastic recurrent neural networks II

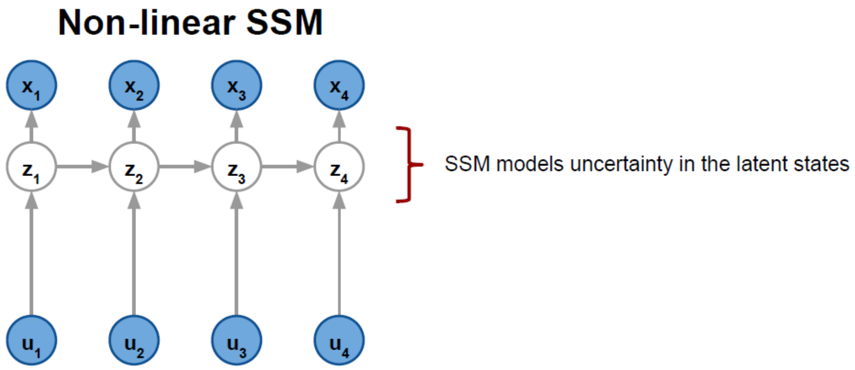


Figure: State-space model [3]

Stochastic recurrent neural networks III

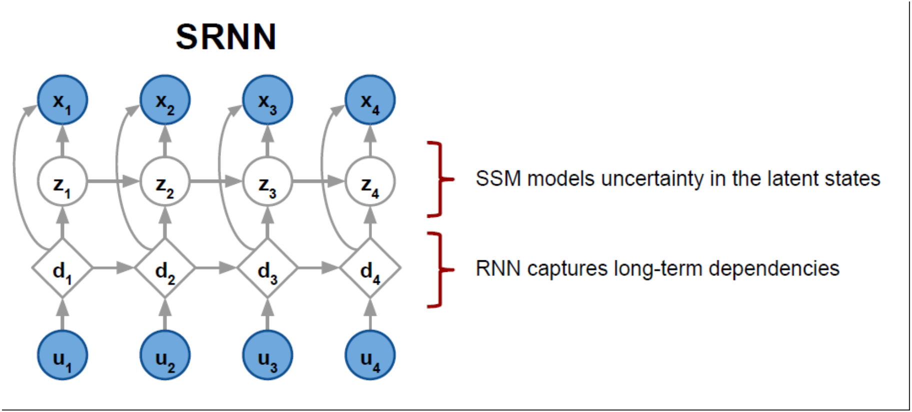


Figure: Merged RNN and state-space model [3]

Differential equations

Infer solutions to partial differential equations, and obtain physics-informed surrogate models [4]

- Neural networks can represent an arbitrary functions when given appropriate weights.
- Therefore it can approximate any arbitrary function that represents a solution of a differential equation: $u = NN(x)$
- Assume that we are given a differential equation with boundary conditions.
- We can also find du/dx, d^2u/dx^2 through back-propagation.
- The goal is to minimize the mean square error loss formed by differential equation and boundary conditions using automated differentiation.

Gaussian processes

- Each data points is a random variable generated from multivariate normal distribution
- The relationship between random variables determines the shape of the latent function.
- Advantages:
  - Regression and prediction with confidence intervals [5]
  - Learning the parameters of the state space models or differential equations [6]
  - Time series where data is not uniformly sampled.
  - Allow for Bayesian optimization

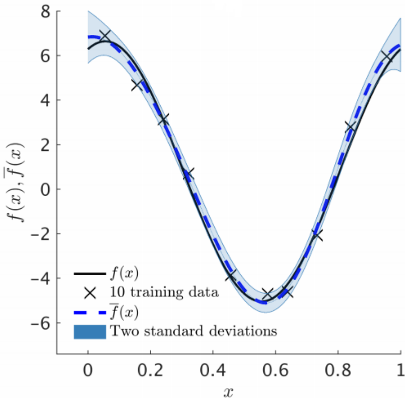


Figure: Gaussian process regression example [6]

State-space models in bioreactors

- State-space models: $dC_i/dt = \mu_i(t)C_1$,
  C_i and μ_i are the concentrations and specific rates of the i^{th} species, respectively, and i represents viable cell density (Xv), concentration of glucose (GLC), lactate (LAC), glutamine (GLN), glutamate (GLU) and ammonia (NH4), and osmolality (Osm), and titer.
- μ_i is estimated based on ML
- C_i is estimated using Extended Kalman filter

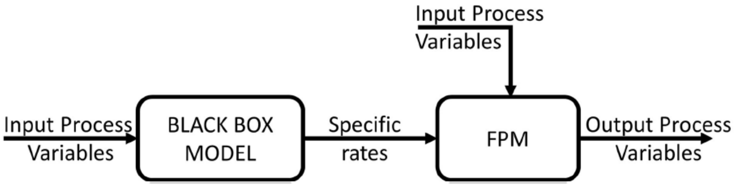


Figure: Hybrid state-space model (FPM is first-principle model) [7]

Suggestion for research I

- Design a simulator that will allow us to simulate the bioprocess and bioreactor based on mechanistic or hybrid model [7]. This is important for:
  - digital twin
  - generating data for testing algorithms
- Merging mechanistic and ML models in this field has just started
  - there is great research opportunity to be first to apply some of these ML approaches on data from bioreactor.
- Gaussian processes

Suggestion for research II

- Sequential decision making: Incorporating human knowledge into:
  - Reward
  - Policy and action selection
- Pre-training or intelligent initialization of the parameters of the ML model
  - Transfer learning

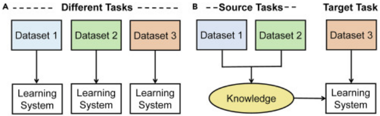


Figure: Transfer learning [8]

  - Meta learning
    - learning from other processes
    - from our data collected using different sensors and in different ways

References I

[1] von Rueden L, Mayer S, Beckh K, Georgiev B, Giesselbach S, Heese R, et al. Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems. 2020.

[2] Jia X. ML applications in scientific domains; 2020. Available from: https://drive.google.com/file/d/1WklAX7vVQ6ehKvCQQCb13Lw_T7WG8aqY/view.

[3] Fraccaro M, Sønderby SK, Paquet U, Winther O. Sequential Neural Models with Stochastic Layers. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16. Red Hook, NY, USA: Curran Associates Inc.; 2016. p. 2207–2215.

[4] Raissi M, Perdikaris P, Karniadakis GE. Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations; 2017.

[5] Swain SKLA P. Inferring time derivatives including cell growth rates using Gaussian processes. Nat Commun. 2016;7.

References II

[6] Raissi M, Perdikaris P, Karniadakis GE.
Machine learning of linear differential equations using Gaussian processes.
Journal of Computational Physics. 2017 Nov;348:683–693.
Available from: http://dx.doi.org/10.1016/j.jcp.2017.07.050.

[7] Narayanan H, Behle L, Luna MF, Sokolov M, Guillén-Gosálbez G, Morbidelli M, et al.
Hybrid-EKF: Hybrid model coupled with extended Kalman filter for real-time monitoring and control of mammalian cell culture.
Biotechnology and Bioengineering. 2020;117(9):2703–2714.
Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.27437.

[8] Deng C, Ji X, Rainey C, Zhang J, Lu W.
Integrating Machine Learning with Human Knowledge.
iScience. 2020;23(11):101656.
Available from: http://www.sciencedirect.com/science/article/pii/S2589004220308488.