Algo Epi Reading Group

Systems biology informed deep learning for inferring parameters and hidden dynamics

Alireza Yazdani, Lu Lu, Maziar Raissi, George Em Karniadakis

January 14, 2022

Contents

IOWA

Introduction

Background

- Traditionally, systems biological processes are modelled using a system of Ordinary Differential Equations (ODEs).
- With data available, the actual challenge lies in computing the estimated values of the parameters of the ODEs.
- In biological reaction networks, experimental data are insufficient considering the size of the model, which results in parameters that are non-identifiable or only identifiable within confidence intervals.
- Some models are highly sensitive to slight change in parameter values.

Contributions

- It uses a Neural Network Model to infer the hidden dynamics of experimentally unobserved species as well as the unknown parameters in the system of equations.
- The model adds constraints to the optimization algorithm, which makes the method robust to measurement noise and few scattered observations.
- The algorithm is computationally scalable and feasible, and its output is interpretable even though it depends on a high-dimensional parameter space.

Problem

Preliminaries

- Systems biological processes can be modeled by a system of ODEs of the form:

$$\frac{d\mathbf{x}}{dt} = \boldsymbol{f}(\mathbf{x}, t; \mathbf{p}), \tag{1a}$$

$$\mathbf{x}(T_0) = \mathbf{x}_0, \tag{1b}$$

$$\mathbf{y} = \boldsymbol{h}(\mathbf{x}) + \boldsymbol{\epsilon}(t), \quad \boldsymbol{\epsilon}(t) \sim \mathcal{N}(0, \sigma^2), \tag{1c}$$

$x = (x_1, x_2, \ldots, x_S)$ is the concentration of S species.

$p = (p_1, p_2, \ldots, p_K)$ are the K parameters of the model.

$y = (y_1, y_2, \ldots, y_M)$ are the M measurable signals. ($M \le S$)

h() could be any function (in this case, is considered linear).

Preliminaries cont.

- Є is just Gaussian noise.

If h() is a linear function, the system could look like:

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_M \end{pmatrix} = \begin{pmatrix} x_{s_1} \\ x_{s_2} \\ \cdots \\ x_{s_M} \end{pmatrix} + \begin{pmatrix} \epsilon_{s_1} \\ \epsilon_{s_2} \\ \cdots \\ \epsilon_{s_M} \end{pmatrix}, \tag{2}$$

The SIR System of ODEs

Considering the system of ODEs of the simple SIR model, the equations can be written as:

- 3 states (Susceptible (S), Infected (I) and Recovered (R)). So, here S= 3
- So, β and γ are the parameters of the ODE. p= (β, γ).

$$\frac{dS}{dt} = -\frac{\beta S I}{N}$$

$$\frac{dI}{dt} = \frac{\beta S I}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

IOWA

Cont.

- If β = 0.6 and γ = 0.4, then x may look something like this:

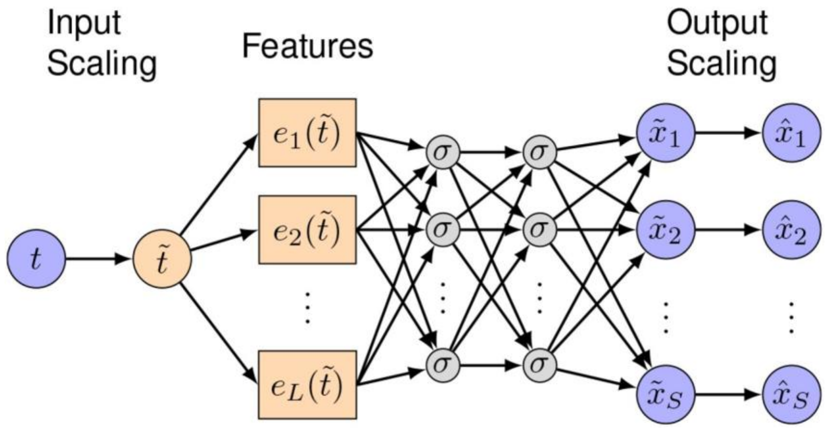
Given a population of 100 people,

Time Period	Susceptible	Infected	Recovered
0	999	1	0
1	999	1	0
2	998	2	0
3	996	2	1

The Model

Neural Network Structure

- **Input:** time t, **Output:** Vector of state variables $\hat{x}(t,\theta)$
- **Input Scaling Layer:** $\bar{t} = \frac{t}{T}$.
- **Feature layer:** $\bar{t}\ == \left(e_1(\bar{t}), e_2(\bar{t}), \dots, e_\mathrm{L}(\bar{t})\right)$.
- **Output Scaling Layer:** Similar to input scaling.



Constrain NN to satisfy ODE

- If we have measurements $y_1, y_2, \ldots, y_M$ at $t_1, t_2, \ldots, t_{Ndata}$ and we enforce the NN to satisfy the ODE system at time points $\tau_1, \tau_2, \ldots, \tau_{N_{ode}}$, then the total loss is defined as follows:

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{p}) = \mathcal{L}^{data}(\boldsymbol{\theta}) + \mathcal{L}^{ode}(\boldsymbol{\theta}, \mathbf{p}) + \mathcal{L}^{aux}(\boldsymbol{\theta}),$$

- The loss function is minimized by the Adam optimizer, given by:

$$\boldsymbol{\theta}^*, \mathbf{p}^* = \arg\min_{\boldsymbol{\theta}, \mathbf{p}} \mathcal{L}(\boldsymbol{\theta}, \mathbf{p}).$$

Parts of the Loss Function

- **First Part:**

$$\mathcal{L}^{data}(\boldsymbol{\theta}) = \sum_{m=1}^{M} w_m^{data} \mathcal{L}_m^{data} = \sum_{m=1}^{M} w_m^{data} \left[\frac{1}{N^{data}} \sum_{n=1}^{N^{data}} \left(y_m(t_n) - \hat{x}_{s_m}(t_n; \boldsymbol{\theta}) \right)^2 \right], \quad (4)$$

It is associated with M sets of observations **y** given in equation 1(c).

It is a supervised loss function.

Cont..

- **Second Part:**

$$\mathcal{L}^{ode}(\boldsymbol{\theta}, \mathbf{p}) = \sum_{s=1}^{S} w_s^{ode} \mathcal{L}_s^{ode} = \sum_{s=1}^{S} w_s^{ode} \left[\frac{1}{N^{ode}} \sum_{n=1}^{N^{ode}} \left(\frac{d\hat{x}_s}{dt}|_{\tau_n} - f_s\left(\hat{x}_s(\tau_n;\boldsymbol{\theta}), \tau_n; \mathbf{p}\right) \right)^2 \right], \tag{5}$$

This loss enforces the structure imposed by the system of ODEs given in equation 1(a).

This is an unsupervised loss function.

IOWA

Cont...

- **Third Part:**

$$\mathcal{L}^{aux}(\boldsymbol{\theta}) = \sum_{s=1}^{S} w_s^{aux} \mathcal{L}_s^{aux} = \sum_{s=1}^{S} w_s^{aux} \frac{(x_s(T_0) - \hat{x}_s(T_0; \boldsymbol{\theta}))^2 + (x_s(T_1) - \hat{x}_s(T_1; \boldsymbol{\theta}))^2}{2}. \quad (6)$$

This is an additional source of information for the system identification and involves 2 time instants T0 and T1.

It is essentially a component of the data loss; however, we prefer to separate this loss from the data loss, as in the auxiliary loss data are given for all state variables at these 2 time instants.

This is also supervised loss.

Training

Step 1 Considering that supervised training is usually easier than unsupervised training, we first train the network using the two supervised losses $\mathcal{L}^{data}$ and $\mathcal{L}^{aux}$ for some iterations, such that the network can quickly match the observed data points.

Step 2 We further train the network using all the three losses.

Experiments

The Yeast Glycolysis Model

The System of ODEs

$$\frac{dS_1}{dt} = J_0 - \frac{k_1 S_1 S_6}{1 + (S_6/K_1)^q}, \tag{S4a}$$

$$\frac{dS_2}{dt} = 2\frac{k_1 S_1 S_6}{1 + (S_6/K_1)^q} - k_2 S_2 (N - S_5) - k_6 S_2 S_5, \tag{S4b}$$

$$\frac{dS_3}{dt} = k_2 S_2 (N - S_5) - k_3 S_3 (A - S_6), \tag{S4c}$$

$$\frac{dS_4}{dt} = k_3 S_3 (A - S_6) - k_4 S_4 S_5 - \kappa (S_4 - S_7), \tag{S4d}$$

$$\frac{dS_5}{dt} = k_2 S_2 (N - S_5) - k_4 S_4 S_5 - k_6 S_2 S_5, \tag{S4e}$$

$$\frac{dS_6}{dt} = -2\frac{k_1 S_1 S_6}{1 + (S_6/K_1)^q} + 2k_3 S_3 (A - S_6) - k_5 S_6, \tag{S4f}$$

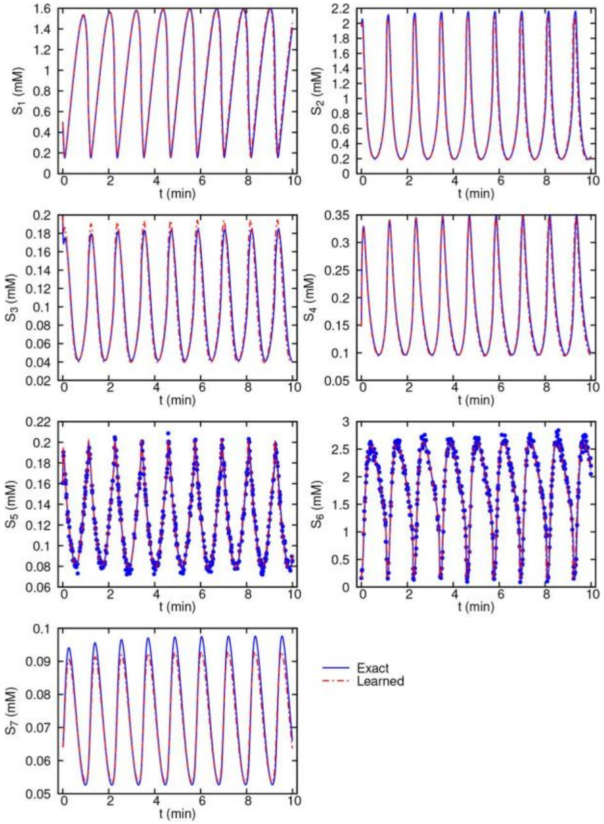
$$\frac{dS_7}{dt} = \psi \kappa (S_4 - S_7) - k S_7, \tag{S4g}$$

IOWA

Steps

- Observation data is corrupted with Gaussian noise.
- Initially, dynamics are inferred using noiseless observations on 2 species S6 and S5 only.
- Sampling of data points between 1-10 mins at random is used to train the NN.
- Then, inferred dynamics of S5 and S6 is compared with noisy data.
- The parameters p of the ODE are inferred by backpropagation to estimate NN parameters θ.

Results



Parameter	Target value	Inferred value (Noiseless observations)	Inferred value (Noisy observations)	Standard deviation
J_0	2.5	2.50	2.49	0.18
k_1	100	99.9	86.1	62.0
k_2	6	6.01	4.55	21.3
k_3	16	15.9	14.0	21.9
k_4	100	100.1	97.1	103.6
k_5	1.28	1.28	1.24	0.25
k_6	12	12.0	12.7	5.1
k	1.8	1.79	1.55	4.34
κ	13	13.0	13.4	25.9
q	4	4.00	4.07	0.27
K_1	0.52	0.520	0.550	0.091
ψ	0.1	0.0994	0.0823	0.317
N	1	0.999	1.29	2.94
A	4	4.01	4.25	2.28

IOWA

Discussions

Points to be Noted

- We are able to infer the unknown parameters of the system of ODEs once the neural network is trained
- In this model we can use a minimalistic amount of data on a few observables to infer the dynamics and the unknown parameters.
- The goal in this work was not to do systematic identifiability analysis, but rather to use identifiability analysis to explain some of their findings.

Thank You

→ https://github.com/alirezayazdani1/SBINNs