Personalized Computational Causal Modeling with Neural Networks

Note 2: Population Model Construction

February 24, 2025

1 Definitions

- **Sample**: All data of a patient.
- **Data point**: The vector (A_β, τ, N, C) of a patient at a certain date.

2 Data Processing

2.1 Data Reduction

To process data efficiently, we first apply data reduction:

1. Using dataset: ANDI Org. and CSF Biomarker
2. Calculating each patient's age at the drwdate, keeping one decimal place
3. Combining data of the same patient and age to one data, i.e. collecting A_β, τ, N and C data of same patient and age together.
4. Similar to process in [1], removing samples which do not provide at least two measurements for any one of the four biomarkers.
5. Calculating each biomarkers' population 5 quartile x_5 and 95 quartile x_{95}, conducting normalize based on 5 and 95 percentile:
$$\tilde{x} = \frac{x - x_5}{x_{95} - x_5} \tag{1}$$
where $\tilde{x}$ denotes the normalized data.

3 Constructing Population A_β Model

3.1 Model Structure

We assume that data of each patients' A_β amount increases following a dynamical system:

$$\frac{dA_\beta}{ds} = f_1(x) \tag{2}$$

where $f(x)$ is a 1-32-32-1 neural network with activation function ReLU of all neurons.

3.2 Training

3.2.1 DPS

Similar to the process in article [1], Page 9, equation(3), we apply linear regression to each patient's age:

$$\mathbf{s_i} = a_i \mathbf{t_i} + b_i \tag{3}$$

where $\mathbf{t}$ denotes patient i's age vector (age of each data points) and $\mathbf{s}$ denotes the **DPS (Disease progression scores)**. Here we set $a_i = 0.1$ and $b_i = -10$ as the initial value. In the training procedure, a_i and b_i will be optimized to each samples. **The initial value a_i and b_i may make a great influence to the model behavior, so we would discuss their influence in dynamical systems in the future.**

3.2.2 Optimization

Integrating f_1 with Forward Eular Method:

$$A_\beta(s_j) = A_\beta(s_{j-1}) + f_i(A_\beta(s_{j-1}))(s_j - s_{j-1}) \tag{4}$$

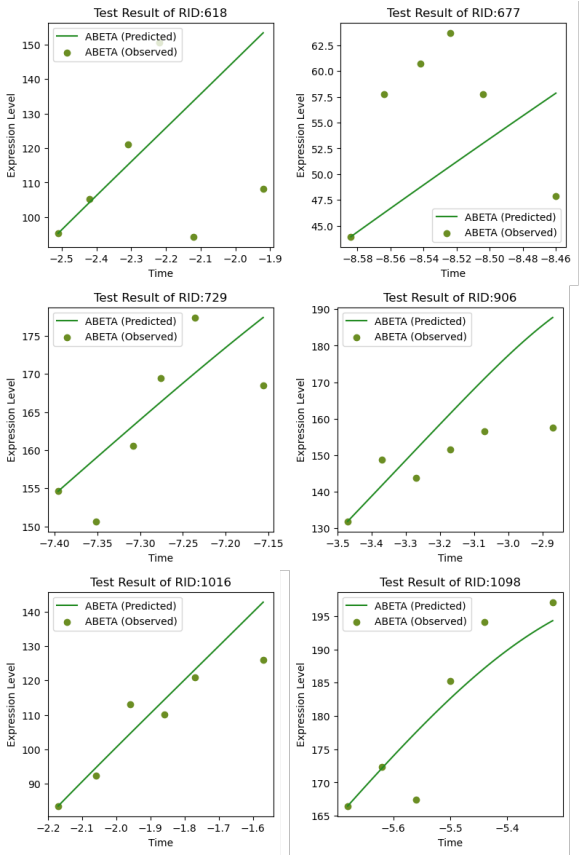
here j denotes j-th time point, sample ids i were omitted.
We will get the prediction vector $\mathbf{A}_{\beta i}$ of each sample. We conduct gradient descent on a_i and b_i, then move on to the next sample. After all samples are processed, we calculate the average loss of each sample, then apply gradient descent to the neural network's parameters.

Neural Network initialization	Kaiming
Optimizer of Neural Network	Adam
Optimizer of a and b	Adam
Learning rate of 2 optimizers	0.01
Epoch	200

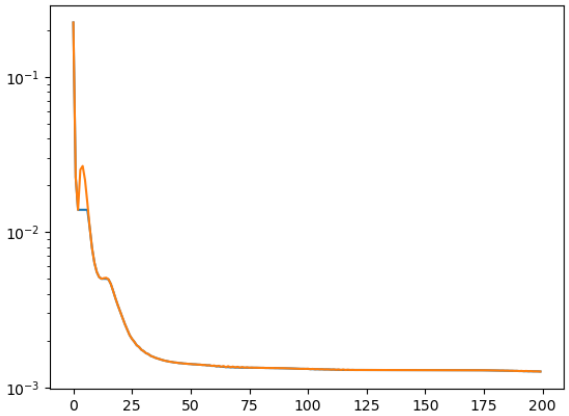
Table 1: Learning settings

3.2.3 Results

Here we selected 6 samples in 1a to compare real data and prediction results.



(a) Selected prediction results of A_β



(b) Loss Curve of A_β

Figure 1: Prediction results and loss curve for A_β

The loss curve is also shown in 1b. Remind that as we used all samples to train the neural network, i.e. we used Gradient Descent but not SGD, so the loss curve is smooth and monotonous in most time.

As shown in [1], Page 10, equation(6), Zheng et al used a system of polynomials to describe the Alzhemer's Disease dynamical system. The population parameters of the system is shown in Page 4, Table 1. Here in the Figure 2 we would compare our neural network model f_1 with the polynomial model P_1.

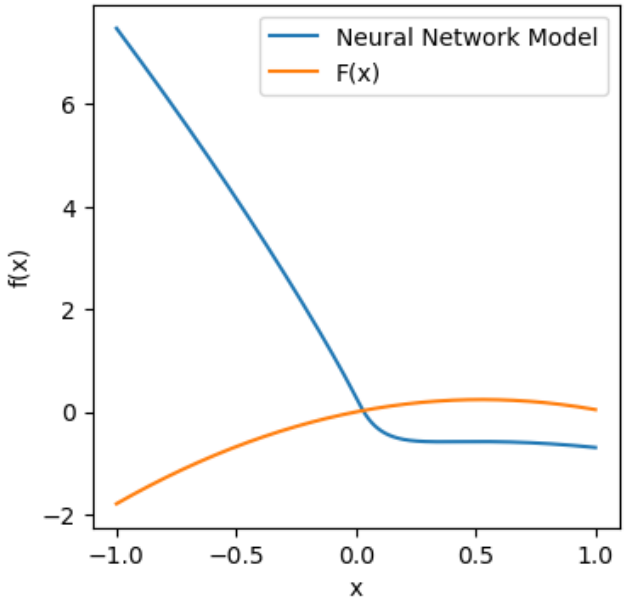


Figure 2: Comparison between Polynomial and Neural Network for A_β

4 Constructing Population τ Model

4.1 Model Construction

The model construction process for τ is similar to A_β, the only distinction is that τ's derivative has 2 inputs: A_β and τ:

$$\frac{d\tau}{ds} = f_2(A_\beta, \tau) \tag{5}$$

To predict τ, we first check if all time points s_j have corresponding $_{\beta j}$ value. If there is, we would use this true value as input. If not, we would use the nearest backward true $A_{\beta j'}$ value as input for Equation 2, to predict $_{\beta j}$. Therefore, our input A_β has the most authentic composition. There are more details in the program, but we would not discuss here.

The selected prediction results of τ model are listed in Figure 3.

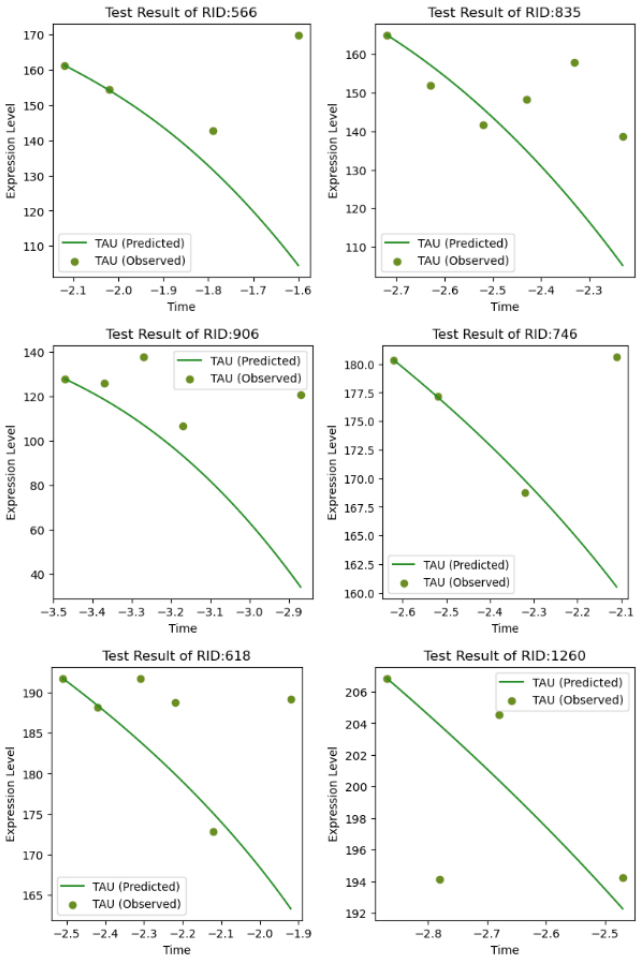


Figure 3: Selected prediction results of τ

Similarly, we would compare the neural network function f_2 with P_2:

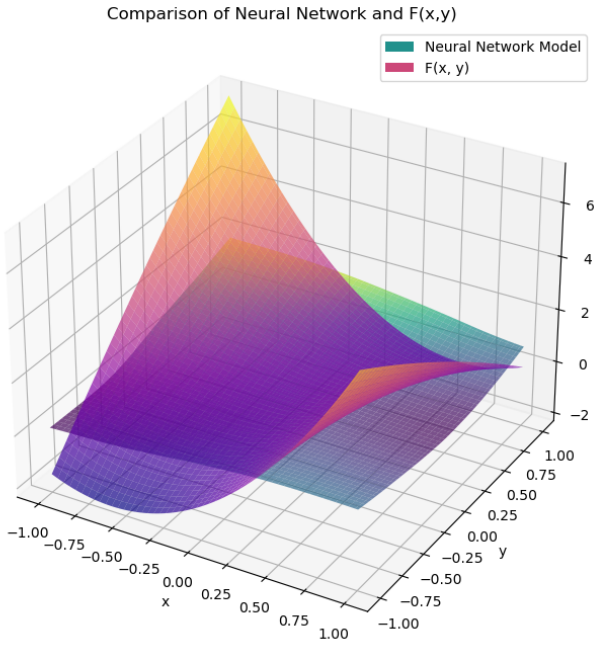


Figure 4: Comparison between Polynomial and Neural Network for τ

The loss curve is similar to case in A_β, so we would not show it here. Remind that training settings is same

as case in A_β, and DPS parameters a_i and b_i were trained independently.

5 Constructing Population N and C Model

The model of N and C is:

$$\frac{dN}{ds} = f_3(\tau, N) \tag{6}$$

$$\frac{dC}{ds} = f_4(N, C) \tag{7}$$

We only put prediction results and function curve here.

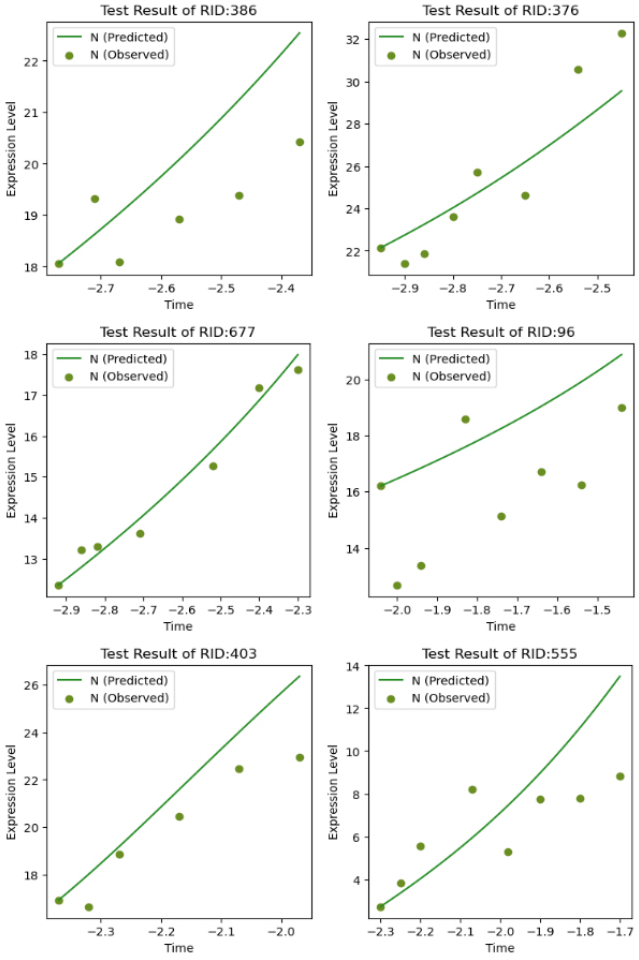


Figure 5: Selected Prediction Results of N

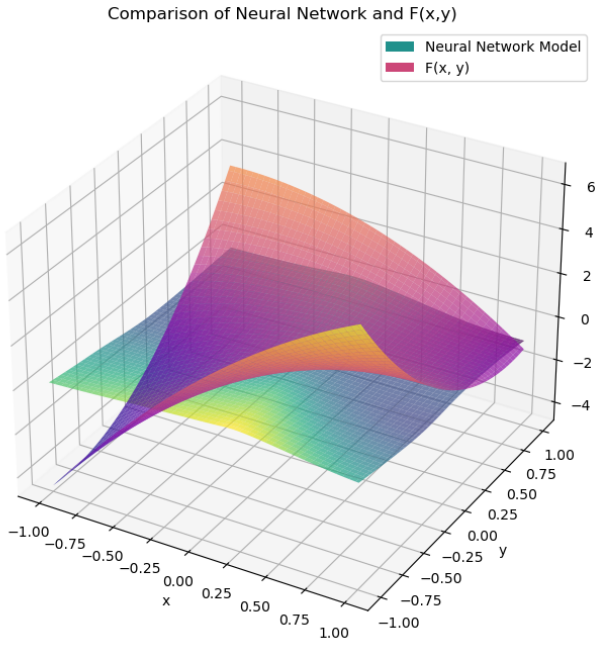


Figure 6: Comparison between Polynomial and Neural Network for N

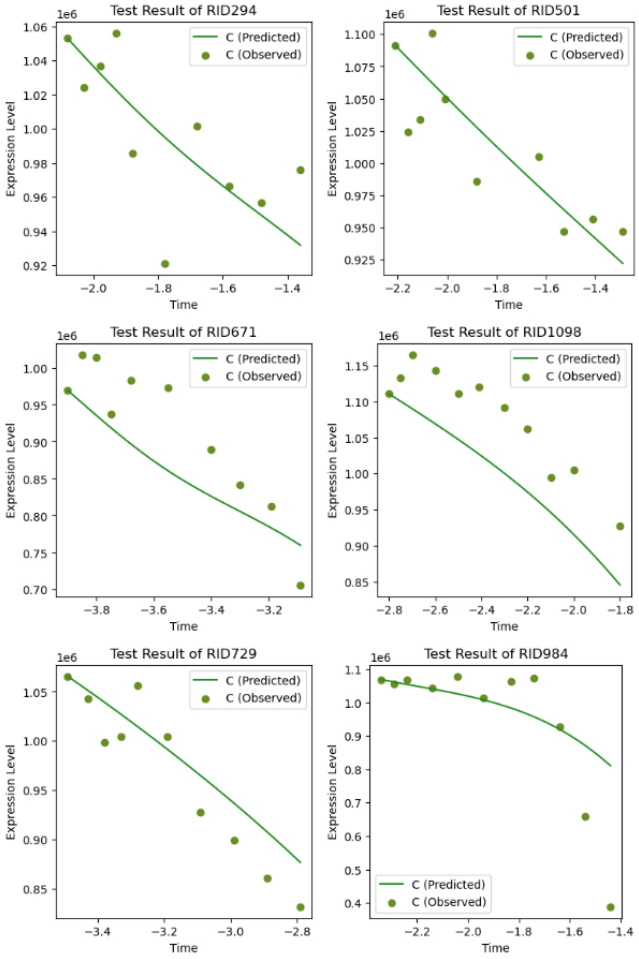


Figure 7: Selected Prediction Results of C

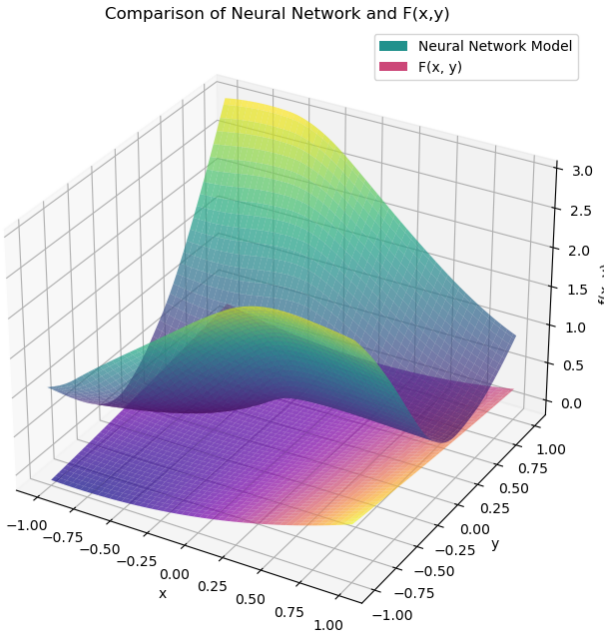


Figure 8: Comparison between Polynomial and Neural Network for C

6 Prospects for Future Research

The upcoming work may focus on:

1. Conducting sensitivity analysis and constructing personalized models
2. Compare our approach with NeuralODE
3. Change loss function to precise numerical derivatives. Improve model's precision, especially integration precision
4. Analyze Dynamical System properties, such as Attractors, Lyapunov Stability
5. Reduce model scale

References

[1] H. Zheng, J. R. Petrella, P. M. Doraiswamy, et al. Data-driven causal model discovery and personalized prediction in alzheimer's disease. *npj Digital Medicine*, 5(1):137, 2022.