Personalized Computational Causal Modeling with Neural Networks

Note 1: Data Processing and Training

2025/01/18

1 Data Processing

i)

In the original paper's equation (1):

$$\begin{cases} \dfrac{dA_\beta}{dt} = \lambda_{A_\beta} A_\beta (K_{A_\beta} - A_\beta) \\ \dfrac{d\tau_\rho}{dt} = \lambda_{\tau_\rho A_\beta} A_\beta + \lambda_{\tau_\rho} \tau_\rho (K_{\tau_\rho} - \tau_\rho) \\ \dfrac{d\tau_o}{dt} = \lambda_{\tau_o AS} \\ \dfrac{dN}{dt} = \lambda_{N\tau_o} \tau_o + \lambda_{N\tau_\rho} \tau_\rho + \lambda_N N (K_N - N) \\ \dfrac{dC}{dt} = \lambda_{CN} N + \lambda_C C (K_C - C) \end{cases} \tag{1}$$

N and C are data from sheet2: ADNI Org. After comparison, there are many mismatches between EXAMDATE and the DRWDTE of the three protein markers (hereinafter referred to as markers). For example, patient RID4's N and C were sampled from:.
2005/11/8, 2006/5/2, 2006/11/14, 2007/5/14, 2008/5/8, 2008/11/18.
But his markers were sampled from:
2005/11/22, 2005/11/22, 2005/11/22, 2006/11/28, 2006/11/28

My idea is that, according to equation (1), the dynamics of markers are independent of N and C and are not affected by them. Therefore, we can first establish the marker model, then simulate the temporal changes in marker content for each patient (including historical data), and then train models corresponding to equations 4 and 5 based on these simulated data.

ii)

Since the Neural ODE package requires strictly monotonic timelines, for the large number of duplicate samples in sheet3 (samples with the same RID and DRWDTE but different specific data), we merged them by calculating their average values. For example:

RID	VISCODE	BATCH	KIT	STDS	DRWDTE	RUNDATE	ABETA	TAU
3	bl	UPENNBIOMK	170470	167082	2005/9/12	2007/10/26	131	68
3	bl	UPENNBIOMK2	190841	181286	2005/9/12	2008/11/12	132	54.9
3	bl	MEDIAN	ALL	ALL	2005/9/12	2016/3/9	131	61.4

Patient RID3 has three samples with identical DRWDTE but different marker data, so we calculate their average and merge them into one.

Meanwhile, there are many samples in the dataset with only one time point (i.e., a patient with only one data entry). Such data cannot be processed by Neural ODE, and we cannot combine data from different patients into one timeline, so these must be discarded.

After statistics, only 624 samples meet our requirements, and only 154 are high-quality samples (samples with at least three time points).

iii)

The data underwent the same standardization process as in the original paper, namely first calculating the 5th and 95th percentiles of each quantity, then normalizing them according to these percentiles.

Specifically, time was also normalized: using October 22, 2007, as the baseline, calculating the time difference between each individual's baseline date and sampling date, then normalizing.

2 Training

i)

The following neural network structures were attempted:

- 3-n-3 neural network
- 3-n-n-3 neural network, which contains many quadratic terms. I hoped to create a small-scale neural network that, if performing well, would allow us to obtain all its parameters and analytical expressions, then transform it to approach the format of equation (1) and assign meaning to the parameters
- 3-1024-3 neural network, i.e., a shallow but wide neural network

ii)

The following combinations of activation functions were attempted:

- Relu
- Elu
- $Relu^2$
- Elu^2
- Sigmoid

iii)

At the beginning of each epoch, a random patient is selected, and all their data is used for training.

During training, the current neural network's performance is evaluated on random 10 samples (patients). If it achieves historical best performance, the current model is saved as the best model. The final returned model is the best model.

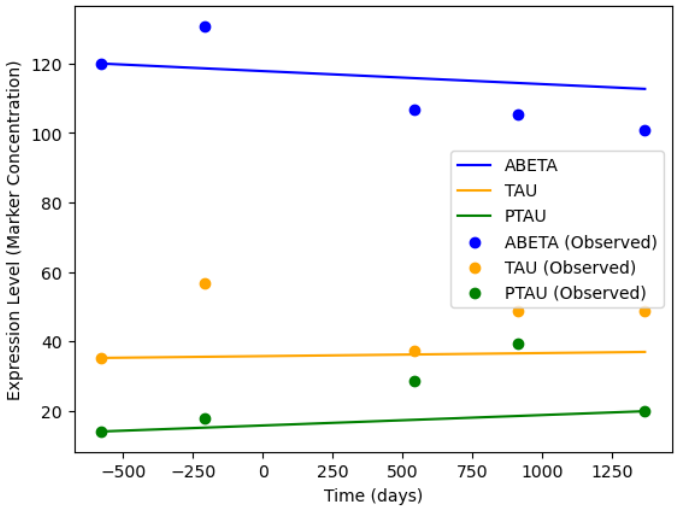
iv)

L2 regularization technique and Xavier initialization technique were used during training.

v)

In the above attempts, the neural network did not perform well. Specifically, the simulated results were almost straight lines, meaning the neural network's output was approximately constant.

The prediction results are shown in the following figure (the data in the figure has been inverse-normalized, and the figure shows a randomly selected sample for testing):



vi)

Since the neural network is a mapping from R^3 to R^3 and acts as a derivative function in equation (1), we can view the neural network as a tangent vector field on $\hat{R3}$ (samples correspond to any trajectory), which can be visualized:

This figure shows that the neural network is indeed a constant mapping.