UTHealth®
The University of Texas
Health Science Center at Houston
School of Biomedical Informatics

Models of Neural Ordinary Differential Equations in Deep Learning Prediction Models Based on Irregularly-sampled Time Series Electronic Health Record Data

Bingyu Mao
BMI 7302: Theories and Frameworks for Biomedical Informatics Research
The School of Biomedical Informatics |The University of Texas Health Science Center at Houston

Please contact the first author via email: Bingyu.Mao@uth.tmc.edu

INTRODUCTION

Analyzing EHR (electronic health record) data allows us to improve clinical decision support and predict clinical processes for specific conditions. Nevertheless, time-series EHR data are always incomplete and irregularly sampled (Figure 1).

The goal of this study is to find proper neural ODE-based models for irregularly-sampled time series EHR data analysis.

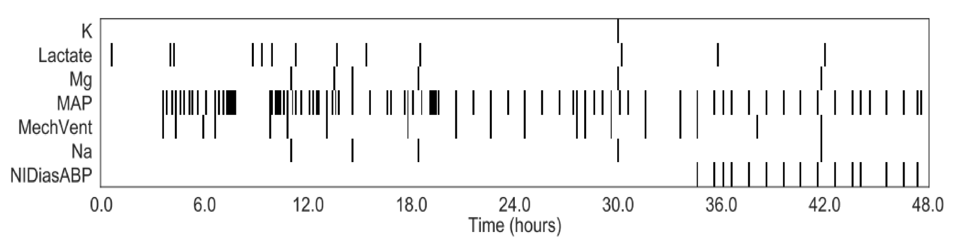

Figure 1: An example of time-series EHR data (Rubanova et al., 2019)

ORDINARY DIFFERENTIAL EQUATION-BASED MODELS

1. **Neural ODE**: To make neural networks continuously update hidden states, Chen et al. (2018) suggest creating a type of neural network called neural ODE (ordinary differential equation). However, the trajectory of an ODE is only related to its initial condition and equation, which means that the model's trajectory cannot be adjusted based on observations.
2. **Neural CDE (Controlled differential equation)**: As a result of incorporating observational information into the model, neural CDE is able to adjust the trajectory of the model as a result of observed data being fed into the model.
3. **ODE2VAE**: Combines the VAE (variational autoencoder) model with the neural ODE, allowing the dynamic delay of the decomposed ODE to be modeled as location and timing.
4. **Latent ODE**: Defines an evolution process based on the deterministic evolution of the initial latent state in time, on which time series are generated.

STOCHASTIC DIFFERENTIAL EQUATION-BASED MODELS

1. **Neural Jump SDE**: An extension of the neural ODE framework to include discontinuities for models of hybrid systems to model the effects of sudden events.
2. **SSM-SDE (State-space model)**: A mixed model of statistical and physiological which combined SDE and state-space model for insulin-glucose dynamics to produce long-term predictions.
3. **Adjoint SDE**: Calculating the gradient of the solution of ODEs as SDEs, also allows us to use solvers that are high-order adaptive to perform time-efficient constant memory costs in calculating the gradient.

SUMMARY & DISCUSSION

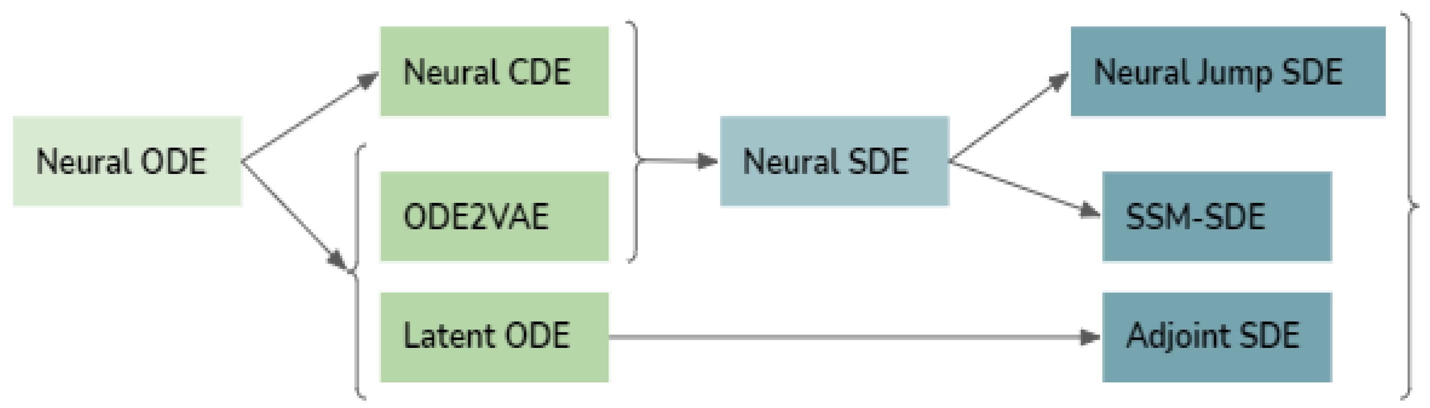

Figure 2:How ODE-based models related to the neural ODE model and also how they are related to each other

Table 1: Summary of models

Model name	Description	Strengths	Weaknesses
Neural ODE	Baseline model.	Updates hidden states continuously.	The model's trajectory cannot be adjusted based on observations.
Neural CDE	Use a natural cubic spline to make the trajectory of the hidden state smoother and use a CDE to allow uncertainty.	Can integrate the afferent data without interrupting the ODE and adjoint backpropagation.	A slower computational speed; too many parameters need to be considered.
ODE2VAE	Let the VAE be an encoder, dividing the image as position and velocity from the input use neural ODE to calculate the dynamics over time.	Suitable for high-dimensional data, especially sequence data.	Unable to handle uncertainty; can't adapt the dynamics to random observed time points.
Latent ODE	An optimized version of ODE2VAE, which directly uses the VAE framework.	Able to naturally manage arbitrary time intervals.	A slower computational speed; too many parameters need to be considered.
Neural Jump SDE	Add jumps into the neural ODE model.	Can be used to model various point processes	Not a real SDE, can't include random trajectories.
SSM-SDE	Use SDE as part of the model because of the nature of the data.	Can improve the prediction of pure statistical or mechanical methods.	Does not model the stochasticity directly.
Adjoint SDE	Augmented from latent ODE and uses SDE as part of a VAE.	Time-efficient and constant-memory computation of gradients.	Poor scalability in memory and time.

CONCLUSION

From the literature, it appears to be worthwhile to use the latent ODE, especially when the dataset is not too large, from the standpoint that the latent ODE solves the problem from a mathematical perspective, even if it is computationally complex requires determining optimal hyperparameters. Alternately, natural cubic spline algorithms can be utilized to smooth the trajectory of the hidden states in a neural CDE model.

ACKNOWLEDGEMENTS

Thank Dr. Susan Fenton for guidance in BMI 7302.

Thank Dr. Degui Zhi, Dr. Ziqian Xie, and Dr. Laila Irasmy Gindy Bekhet from Zhi Group for helpful discussions and suggestions.

REFERENCES

1. Chen, T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in Neural Information Processing Systems 31*, pages 6571–6583. https://arxiv.org/pdf/1806.07366.pdf
2. Jia, J., & Benson, A.R. (2019). Neural Jump Stochastic Differential Equations. *NeurIPS*.
3. Kidger P., Morrill J., Foster J., and Lyons T. (2020). Neural controlled differential equations for irregular time series. *Advances in Neural Information Processing Systems*. https://arxiv.org/pdf/2005.08926.pdf
4. Li, X., Wong, T.L., Chen, R.T., & Duvenaud, D.K. (2020). Scalable Gradients for Stochastic Differential Equations. *ArXiv, abs/2001.01328*.
5. Miller, A.C., Foti, N.J., & Fox, E.B. (2020). Learning Insulin-Glucose Dynamics in the Wild. *ArXiv, abs/2008.02852*.
6. Rubanova, Y., Chen, T. Q., and Duvenaud, D. K. (2019). Latent ODEs for irregularly-sampled time series. *Advances in Neural Information Processing Systems*, pp. 5321–5331. https://arxiv.org/pdf/1907.03907.pdf
7. Yildiz, Ç., Heinonen, M., & Lähdesmäki, H. (2019). ODE2VAE: Deep generative second-order ODEs with Bayesian neural networks. NeurIPS. https://arxiv.org/pdf/1905.10994.pdf