GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series

Edward De Brouwer*, Jaak Simm*, Adam Arany, and Yves Moreau

ESAT-STADIUS
KU Leuven
Leuven, 3001, Belgium
edward.debrouwer,jaak.simm,adam.arany,yves.moreau@esat.kuleuven.be

1 Introduction

Multivariate time series are ubiquitous in various domains of science [2, 6, 7] and much of the methodology for time-series analysis assumes that signals are measured systematically at fixed time intervals. However, much real-world data can be *sporadic* (*i.e.*, the signals are sampled irregularly and not all signals are measured each time). A typical example is patient measurements, which are taken when the patient comes for a visit (*e.g.,* sometimes skipping an appointment) and where not every measurement is taken at every visit. Modeling then becomes challenging as such data violates the main assumptions underlying traditional machine learning methods (such as recurrent neural networks).

Recently, the Neural Ordinary Differential Equation (ODE) model [1] opened the way for a novel, continuous representation of neural networks. As time is intrinsically continuous, this framework is particularly attractive for time-series analysis. It opens the perspective of tackling the issue of irregular sampling in a natural fashion, by integrating the dynamics over whatever time interval needed. Up to now however, such ODE dynamics have been limited to the continuous *generation* of observations (*e.g.*, decoders in variational auto-encoders (VAEs) [4] or normalizing flows [5]).

Instead of the encoder-decoder architecture where the ODE part is decoupled from the input processing, we introduce a tight integration by *interleaving* the ODE and the input processing steps. Conceptually, this allows us to drive the dynamics of the ODE directly by the incoming sporadic inputs. To this end, we propose (1) a continuous time version of the Gated Recurrent Unit and (2) a Bayesian update network that processes the sporadic observations. We combine these two ideas to form the GRU-ODE-Bayes method.

The tight coupling between observation processing and ODE dynamics allows the proposed method to model fine-grained nonlinear dynamical interactions between the variables. As illustrated in Figure 1, GRU-ODE-Bayes can (1) quickly infer the unknown parameters of the underlying stochastic process and (2) learn the correlation between its variables (red arrows in Figure 1). In contrast, the encoder-decoder based method NeuralODE-VAE proposed by [1] captures the

* Equal contribution

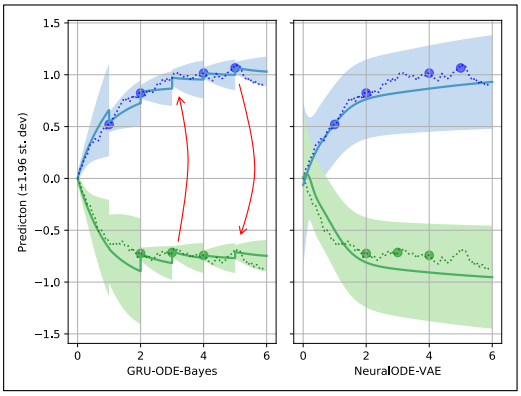

Fig. 1. Comparison of GRU-ODE-Bayes and NeuralODE-VAE on a 2D Ornstein-Uhlenbeck process with highly correlated Wiener processes ($\rho = 0.99$). Dots are the values of the actual underlying process (dotted lines) from which the sporadic observations are obtained. Solid lines and shaded areas are the inferred means and 95% confidence intervals. Note the smaller errors and smaller variance of GRU-ODE-Bayes vs. NeuralODE-VAE. Note also that GRU-ODE-Bayes can infer that a jump in one variable also implies a jump in the other unobserved one (red arrows). Similarly, it also learns the reduction of variance resulting from a new incoming observation.

general structure of the process without being able to recover detailed interactions between the variables.

Our model enjoys important theoretical properties. We frame our analysis in a general way by considering that observations follow the dynamics driven by a stochastic differential equation (SDE). In this paper, we show that GRU-ODE-Bayes can exactly represent the corresponding Fokker-Planck dynamics in the special case of the Ornstein-Uhlenbeck process, as well as in generalized versions of it.

We further perform an empirical evaluation and show that our method outperforms the state of the art on healthcare and climate data. In healthcare, we used electronic health records (EHR) from the MIMIC-III clinical database [3], which contains EHR for more than 60,000 critical care patients. We select a subset of 21,250 patients with sufficient observations and extract 96 different longitudinal real-valued measurements over a period of 48 hours after patient admission. We predicted the next 3 vitals measurements of intensive care patients after 36 hours observations. For the climate application, we used the publicly available United State Historical Climatology Network (USHCN) daily data set ushcn, which contains measurements of 5 climate variables (daily temperatures, precipitation, and snow) over 150 years for 1,218 meteorological stations scattered over the United States. To showcase the capability of our approach, we artificially downsampled the available data and predicted future measurements based on 3 years observations.

References

1. Chen, T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural ordinary differential equations. In: Advances in Neural Information Processing Systems, 2018 (2018)
2. Jensen, A.B., Moseley, P.L., Oprea, T.I., Ellesøe, S.G., Eriksson, R., Schmock, H., Jensen, P.B., Jensen, L.J., Brunak, S.: Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nature communications **5**, 4022 (2014)
3. Johnson, A.E., Pollard, T.J., Shen, L., Li-wei, H.L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Scientific data **3**, 160035 (2016)
4. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
5. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: International Conference on Machine Learning. pp. 1278–1286 (2014)
6. Scargle, J.D.: Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal **263**, 835–853 (1982)
7. Schneider, T.: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. Journal of climate **14**(5), 853–871 (2001)