The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Deep Markov Models for Data Assimilation in Chaotic Dynamical Systems

Calvin Janitra Halim Kazuhiko Kawamoto

Chiba University

Recently, the use of deep learning in data assimilation has been gaining traction. One particular time series model known as deep Markov model has been proposed, along with an inference network that is trained together using variational inference. However, the original paper did not address the full capability of the model in data assimilation problem. Therefore, we aim to evaluate the suitability of a deep Markov model and its inference network against a chaotic dynamical system, which often shows up as a problem in data assimilation. We evaluate the model in various generative conditions. We show that when information about part of the target model is known, the model is able to match the capability of a smoothed unscented Kalman filter, even when there are process and observation noise involved.

1. Introduction

The advancement of computer processors has allowed the simulation of real-world processes, modelled by dynamical systems and numerical models. However, hurdles such as chaos and numerical errors prevented the use of these models for long term forecasting. This gave rise to data assimilation (DA) methods, utilizing both observations and numerical models as inputs to statistical methods to reduce estimation error.

Successes of deep learning have led researchers to focus on combining traditional DA methods with deep learning [Cintra 18]. One specific method trains a neural network-based Gaussian state-space model (GSSM) and an inference network together using variational inference [Krishnan 17].

Nevertheless, [Krishnan 17] didn't evaluate the model on chaotic dynamical systems, the primary target of DA, and the paper didn't address the capability of the model when only part of the generative model is available (namely, transition function and emission function), which indirectly shows the adaptive characteristics of the model. Thus, we aim to assess whether this method can be applied to such systems on various conditions (the availability of transition function or emission function during training).

2. Deep Markov Model

Krishnan et al. proposed a method consisting of a generative model and an inference network trained with variational inference [Krishnan 17]. The method uses evidence lower bound (ELBO) as an objective function, comparing the posterior latents generated by inference network with prior latents, along with maximizing the expected log-likelihood of observations generated by the generative network.

2.1 Generative Model

The generative model takes a sequence of latents and produces the corresponding observations. The model is a

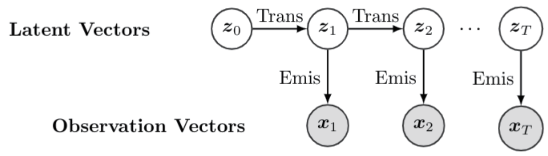

Figure 1: Generative Model (GSSM)

GSSM, whose transition and emission function can be substituted with neural networks, hence the name *deep Markov model* (DMM). The transition function uses a gated transition function (GTU), which is essentially a gated recurrent unit (GRU) [Cho 14] without being conditioned by the observations, similar to the Markovian properties of the latents.

2.2 Inference Network

The inference network takes an observation sequence and infer its corresponding latents. This inference network is structured upon the factorization of the posterior latent distribution, which can be referred to [Krishnan 17]. [Krishnan 17] followed the factorization using a backward recurrent neural network (RNN), which outputs hidden unit for each time step, and then uses a *combiner* function, with the hidden unit and previous latent as input, to output/sample the approximate latent for current time step. As the RNN is propagated from future to the past, [Krishnan 17] uses the notation ST-R, which stands for *STructured-Right*, which we will also adopt. The structure for both generative model and inference network is shown by Figure 1 and Figure 2.

3. Experiments and Results

3.1 Dataset and Evaluation Methods

We train and evaluate the model on Lorenz-96 model [Lorenz 95], which exhibits chaos and is often used on many DA method evaluations. This model also has an atmospheric-like movement, which is similar to real-world atmospherical model. The model is defined by the differen-

Contact: Kazuhiko Kawamoto, Chiba University, 1-33, Yayoicho, Inage Ward, Chiba-shi, Chiba, 263-8522 Japan, +81-43-290-3508, kawa@faculty.chiba-u.jp

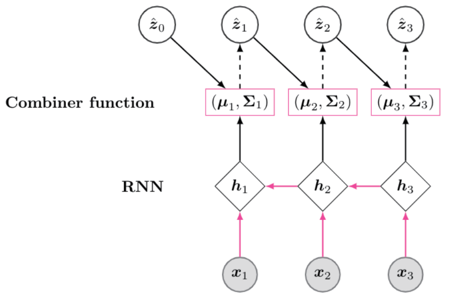

Figure 2: Inference Network (sequence length of 3)

tial equation:

$$\frac{dy^{(i)}}{dt} = (y^{(i+1)} - y^{(i-2)})y^{(i-1)} - y^{(i)} + F \tag{1}$$

where $y^{(i)}$ is the i-th state of the model, $y^{(-1)} = y^{(N-1)}, y^{(0)} = y^{(N)}, y^{(N+1)} = y^{(1)}$ and $\frac{dy^{(i)}}{dt}$ is the derivative of $y^{(i)}$ w.r.t time t. The F here is the forcing constant, which is set into 8.0 to give the model a chaotic movement [Lorenz 95]. We then set the transition and emission function of the generative model into stochastic processes:

$$\textbf{Transtion}: \boldsymbol{z}_t \sim \mathcal{N}(\text{Lorenz96RK4}(\boldsymbol{z}_{t-1}), 5I) \tag{2}$$

$$\textbf{Emission}: \boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{z}_t, 5I) \tag{3}$$

where I is identity matrix, the function Lorenz96RK4(.) is a 4th order Runge-kutta integration function of Lorenz-96 model, with time step difference of 0.01, and $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ denoting a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Both the latent and observation sizes are 20. The training and validation data has 5000 and 500 data each, with sequence length of 25.

We use a smoothed unscented Kalman filter (UKF) as baseline, and take the root-mean-square error (RMSE) between true latents, observations and their reconstructed counterparts as the measurement of the model capability. We also evaluated the model on 4 conditions imposed on the generative model:

1. Condition 1: *fixed transition and emission*
2. Condition 2: *unknown transition and fixed emission*
3. Condition 3: *fixed transition and unknown emission*
4. Condition 4: *unknown transition and emission.*

Here, *fixed* means setting the corresponding function of the generative model into the Equation 2 or 3, and *unknown* means setting the function to be not *known*, which is substituted with neural network that is to be inferred during training process. Also note that the smoothed UKF here is only evaluated on condition 1, as UKF is not an adaptive filter.

3.2 Results

The evaluation result is shown on Figure 3. *ST-R*, *UKF*, *Observations* denote the RMSE (vertical axis) of the DMM (with ST-R), UKF, and observations respectively, and the horizontal axis shows the training epochs. The result shows

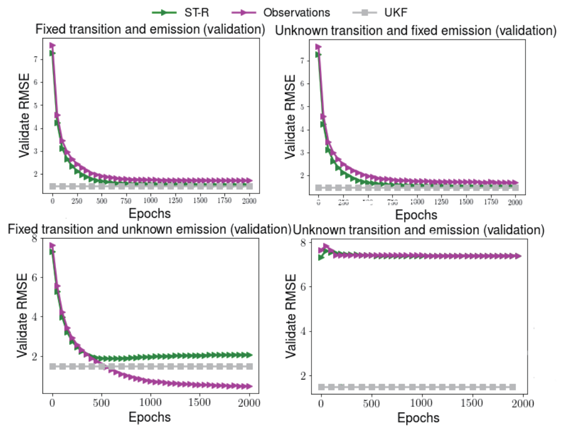

Figure 3: Validation RMSE on 4 conditions

that the RMSE of DMM converges to that of UKF as the training continues on condition 1, 2, and 3. This implies that DMM is capable of state estimation that rivals that of UKF, and is also able to infer part of the generative model when it is not given, showing the capability of DMM as an adaptive filter. However, the model couldn't estimate a system when only the observations are given. We believe that the reason for this lies in the insufficient information of the sampled observations in inferring a generative network that is unknown.

4. Conclusion

We showed that DMM is suitable enough to be used as a DA method, even with a chaotic dynamical system (with the addition of stochasticity) as the target. We plan to experiment further with harder grid problems and improve the model inference capability to surpass current DA methods.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16K00231.

References

[Cintra 18] Cintra, R. S. and Velho, H. F. C.: Data assimilation by artificial neural networks for an atmospheric general circulation model. In *Advanced Applications for Artificial Neural Networks*, ch. 14 (2018)

[Krishnan 17] Krishnan, R. G., Shalit, U. and Sontag, D.: Structured inference networks for nonlinear state space model. In *AAAI* (2017)

[Cho 14] Cho, K., Merrinboer, B., Glehre, ., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on EMNLP*, pp. 1724-1734 (2014)

[Lorenz 95] Lorenz, E.: Predictability: a problem partly solved. In *Seminar on Predictability*, Vol. 1, pp. 1-18 (1995)