Switching Dynamical Systems with Deep Neural Networks

Ojeda, C.[§], Georgiev, B.[†], Cvejoski, K.[†♀], Schuecker, J.[♀], Bauckhage, C.[♀], Sánchez, R.[†‡],

[†]Competence Center Machine Learning Rhine-Ruhr [‡] University of Bonn [♀]Fraunhofer Center for Machine Learning and Fraunhofer IAIS [§]Berlin Center for Machine Learning and TU Berlin

Abstract

The problem of uncovering different dynamical regimes is of pivotal importance in time series analysis. Switching dynamical systems provide a solution for modeling physical phenomena whose time series data exhibit different dynamical modes. In this work we propose a novel variational RNN model for switching dynamics allowing for both non-Markovian and non-linear dynamical behavior between and within dynamic modes. Attention mechanisms are provided to inform the switching distribution. We evaluate our model on synthetic and empirical datasets of diverse nature and successfully uncover different dynamical regimes and predict the switching dynamics.

Motivation

Switching linear dynamical system models aim to capture complex (non-linear) time series behaviour via a collection of so-called dynamical modes, each of which is approximated by a linear model[1, 2]. In short, one assumes that at each time step t there is a corresponding categorical latent state z_t taking one of K different values and following the Markovian transitions

$$z_{t+1} \,|\, z_t \sim \pi_{z_t}, \tag{1}$$

where $\pi_{z_t} \in [0,1]^K$ gives the usual Markov transition probabilities. The classical approach [1, 2] also introduces the continuous latent states $\mathrm{h}_t \in \mathbb{R}^p$ — these follow affine dynamics, with the different modes being indexed by z_t,

$$\mathrm{h}_{t+1} = \mathrm{A}_{z_{t+1}}\mathrm{h}_t + \mathrm{b}_{z_{t+1}} + \mathrm{v}_t, \quad \mathrm{v}_t \sim \mathcal{N}(0, \mathrm{Q}_{z_{t+1}}), \tag{2}$$

where $\mathrm{A}_k, \mathrm{Q}_k$ are matrices of the form $\mathbb{R}^{p\times p}$ whereas $\mathrm{b}_k \in \mathbb{R}^p$ and $k \in (1, \ldots, K)$. At last, the observed data points $\mathrm{x}_t \in \mathbb{R}^d$ are obtained via

$$\mathrm{x}_t = \mathrm{C}_{z_t}\mathrm{h}_t + \mathrm{d}_{z_t} + \mathrm{w}_t, \quad \mathrm{w}_t \sim \mathcal{N}(0, \mathrm{S}_{z_{t+1}}), \tag{3}$$

with $\mathrm{C}_k \in \mathbb{R}^{d\times p}$, $\mathrm{S}_k \in \mathbb{R}^{d\times d}$ and the drift terms $\mathrm{d}_k \in \mathbb{R}^d$.

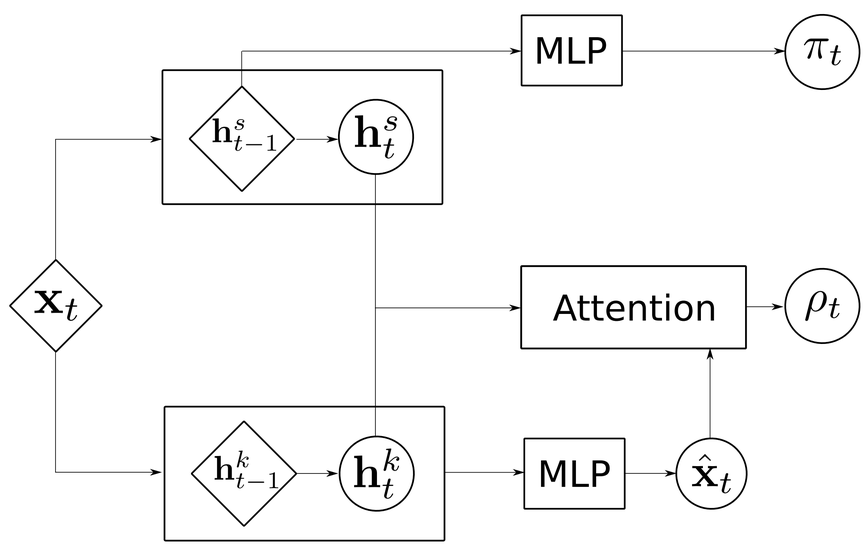
Neural Variational Switching Dynamical Systems

Generative Model

- Switching Dynamics (LSTM) $\quad \mathrm{h}^s_t = f_{\theta_s}\left(\mathrm{x}_t, \mathrm{h}^s_{t-1}\right)$
- Switching Probabilitites $\quad \pi^k_t = \mathrm{softmax}\left[g_{\theta_k}\left(\mathrm{h}^s_{t-1}\right)\right]$
- Switching Prior $\quad p(\mathrm{z}_t) = \prod_{k=1}^{K}\left(\pi^k_t\right)^{z^k_t}$
- Emission Parameters $\quad \left[\boldsymbol{\mu}^k_t, \boldsymbol{\sigma}^k_t\right] = g_{\varphi_k}\left(\mathrm{h}^k_{t-1}\right)$
- Emission Probabilitites $\quad p(\mathrm{x}^k_t \,|\, \mathrm{x}^k_{<t}) = \mathcal{N}(\boldsymbol{\mu}^k_t, \mathrm{diag}\left[(\boldsymbol{\sigma}^k_t)^2\right])$
- Modes Dynamics $\quad \mathrm{h}^k_t = f_{\varphi_k}\left(\mathrm{x}_t, \mathrm{h}^k_{t-1}\right)$

Joint Distribution

$$p(z_{\leq T}, x_{\leq T}) = \prod_{k=1}^{K}\prod_{t=1}^{T}\left(\pi^k_t\, p\left(\mathrm{x}^k_t | \mathrm{x}^k_{<t}\right)\right)^{z^k_t}.$$



Architecture of the NVSDS model. x_t the data is fed into the recurrent network modeling the modes dynamics and switching dynamics (upper part). The representations h obtained by the experts' dynamics are fed into an MLPs parametrizing a Gaussian distribution for the outputs $\hat{x}_t$.

Inference

- Approximate Posterior $\quad q(\mathrm{z}_t|\mathrm{x}_{\leq t}) = \prod_{k=1}^{M}\left(\rho^k_t\right)^{z^k_t}$
- Attention Representation $\quad \mathrm{u}_k = \sigma\left(\mathrm{W}_k\,\mathrm{H}_t + \mathrm{V}_k\,\mathrm{h}^s_t + \mathrm{b}_k\right)$
- Prediction and Hidden States $\quad \mathrm{H}_t = \left([\hat{\mathrm{x}}^1_t, \mathrm{h}^1_t], [\hat{\mathrm{x}}^2_t, \mathrm{h}^2_t], \ldots, [\hat{\mathrm{x}}^K_t, \mathrm{h}^K_t]\right)$
- Modes POsterior Probabilities $\quad \rho^t_k = \mathrm{softmax}\left[\mathrm{u}_k \cdot \mathrm{c}_k\right]$

ELBO

$$\mathcal{L}[q] = \mathbb{E}_{p_D(\mathrm{x})}\left[\sum_{k=1}^{K}\sum_{t=1}^{T}\left\{\rho^k_t\,\log p(\mathrm{x}^k_t|\mathrm{x}^k_{<t}) + \rho^k_t\,\log\left[\frac{\pi^k_t}{\rho^k_t}\right]\right\}\right], \tag{4}$$

Mode Regularization

We enforce dynamical diversity by imposing cost functions to be trained along side the maximization of the lower bound

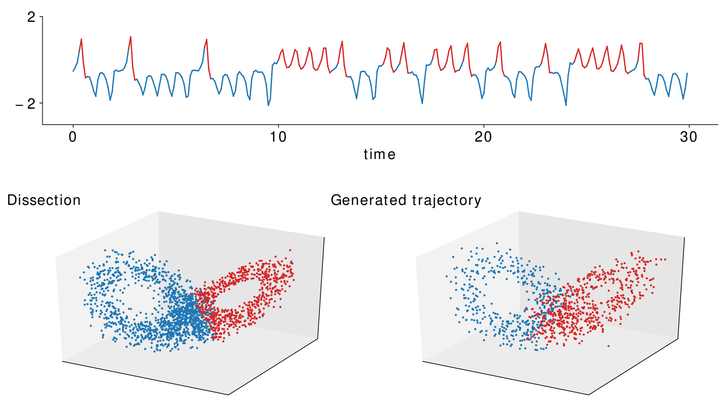
$$\mathcal{H}[\rho] = -\mathbb{E}_{p_D(\mathrm{x})}\sum_{k=1}^{K}\tilde{\rho}_k\,\log\tilde{\rho}_k, \tag{5}$$

where $\tilde{\rho}_k$ is the time-average posterior class probabilities. We maximize then $\mathcal{L}'[q] = \mathcal{L}[q] + \lambda_e\,\mathcal{H}[\rho]$, where λ_e is a hyperparameter.

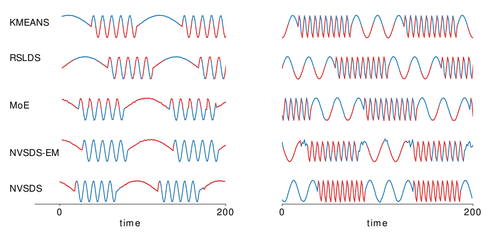
Experiments

Lorenz Attractor It is defined by a coupled system of non-linear equations

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \sigma(y-x),\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= x(\rho - z) - y, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.\end{aligned} \tag{6}$$



Switching Oscillatory Dynamics

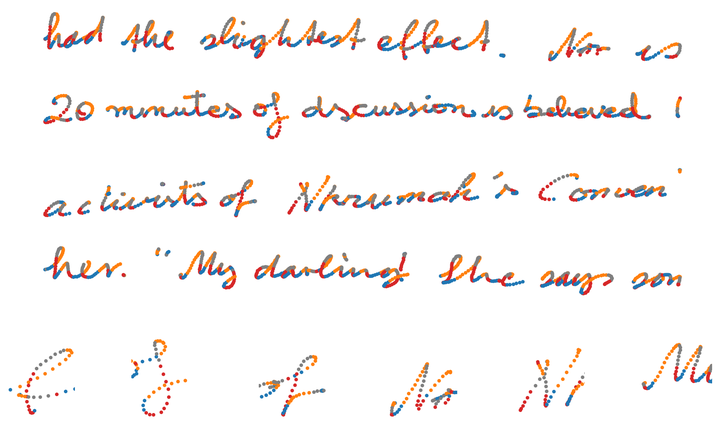


It is defined as:

$$f(t) = H(\cos(\omega_s t) < 0)\cos(\omega_1 t) + [1 - H(\cos(\omega_s t) < 0)]\cos(\omega_2 t) \tag{7}$$

Where H is the Heaviside function and $\omega_s < \omega_2 < \omega_1$
We asses these observations quantitatively by defining a binary target vector $\hat{\rho}$ indicating the dynamical mode which is present at a particular time-step. Evaluating the mean squared error between the predicted ρ and $\hat{\rho}$ yields a dissection error, which we average over multiple trials (different initial conditions). The NVSDS (0.09) and the NVSDS-EM (0.1) clearly outperform R-k-Means (0.26), rSLDS (0.4) and MoE (0.47).

Handwrittig



Dissection of a handwriting signal for the NVSDS model for different sequences. The lower row shows particular letters from the complete sequences for easier comparison.

Conclusion

- In the present work we have provided a neural network solution to the problem of switching dynamical systems (SDS).
- We build upon variational approximate inference for the categorical variables indexing of the dynamical modes.
- We incorporate an attention mechanism for the switching procedure.
- We incorporate an entropy regularizer to improve the detection of the modes.

References

[1] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski, "Bayesian learning and inference in recurrent switching linear dynamical systems," in *Artificial Intelligence and Statistics*, pp. 914–922, 2017.

[2] G. Ackerson and K.-S. Fu, "On state estima-tion in switching environments," *IEEE Transactionson Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.