Switching Dynamical Systems with Deep Neural Networks

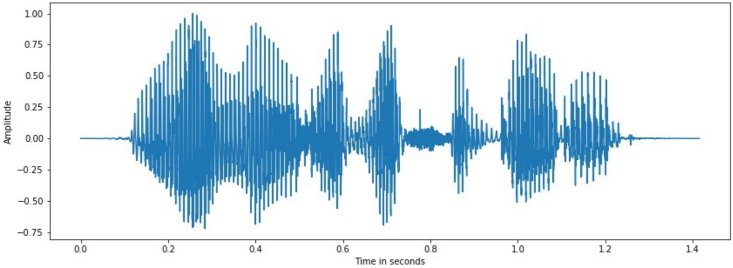
César Ojeda, Ramsés J. Sánchez, Kostadin Cvejoski, Jannis Schücker, Christian Bauckhage and Bogdan Georgiev

TU Berlin, Fraunhofer IAIS, University of Bonn

December 9, 2020

Motivation

Many complex dynamical signals naturally feature an inherent compositional form, in the sense that their data generating process can be decomposed into different dynamical modes.



Switching Dynamical Systems

In Switching Dynamical Systems, 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

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

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

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

Recurrent Switching Linear Dynamical Systems

An important remark is that each categorical state z_{t+1} depends only on the previous one z_t. This seems to limit the influence of the continuous latent variable $\mathbf{h}_t$ on the discrete switch — An augmented model (recurrent SLDS or **rSLDS**) which makes use of the following generation scheme for z_t

$$z_{t+1} \,|\, z_t, \mathbf{h}_t, \{\mathbf{R}_k, \mathbf{r}_k\} \sim \pi_{SB}(\nu_{t+1}), \tag{4}$$

with $\nu_{t+1} = \mathbf{R}_{z_t}\mathbf{h}_t + \mathbf{r}_{z_t}$. Here π_{SB} is a certain stick-breaking distribution, $\mathbf{R}_k \in \mathbb{R}^{K-1\times p}$ captures the recurrent dependencies between z_t and $\mathbf{h}_t$, and $\mathbf{r}_k \in \mathbb{R}^{K-1}$ models the Markovian transitions between consecutive states z_{t+1} and z_t.

Recurrent Neural Networks

Given a sequence $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, RNNs process each $\mathbf{x}_t$ through the update of a hidden state $\mathbf{h}_t$ at each time step $t \in (1, \ldots, T)$. The update is implemented via a deterministic non-linear transition function f_θ thus

$$\mathbf{h}_t = f_\theta\left(\mathbf{h}_{t-1}, \mathbf{x}_t\right), \tag{5}$$

where $\mathbf{h}_t \in \mathbb{R}^p$, $\mathbf{x}_t \in \mathbb{R}^d$ and θ is the parameter set of f. Given the set of hidden states $\mathbf{h}_t$ one can model the observed sequence by approximating its joint probability distribution function as

$$p(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T) = \prod_{t=1}^{T} p(\mathbf{x}_t | \mathbf{x}_{<t}), \quad p(\mathbf{x}_t | \mathbf{x}_{<t}) = g_\varphi(\mathbf{h}_{t-1}), \tag{6}$$

where g with parameter set φ maps $\mathbf{h}_t$ to a probability distribution over outputs, and where $\mathbf{x}_{<t}$ denotes the dependence on the history.

Variational Inference

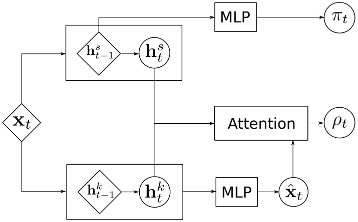
We define an approximate posterior distribution $q(\mathbf{z})$ which is tractable. This distribution is chosen to approximate the unknown true posterior distribution - by minimizing e.g. the Kullback-Leibler divergence $\mathsf{KL}\left[q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})\right]$. The posterior is not explicitly available, we maximize the lower bound to the model's evidence:

$$\mathcal{L}[q] = \mathbb{E}_{q(\mathbf{z})}\left[\log p(\mathbf{x}|\mathbf{z})\right] - \mathsf{KL}\left[q(\mathbf{z})||p(\mathbf{z})\right], \tag{7}$$

where the first term is the averaged log likelihood of the model over the approximate posterior distribution and drives the learning of the data, whereas the second term plays the role of a regularizer.

Neural Variational Switching Dynamical Systems

Generative Model

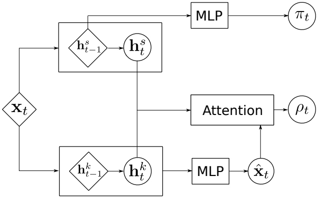


- $\mathbf{h}_t^s = f_{\theta_s}\left(\mathbf{x}_t, \mathbf{h}_{t-1}^s\right)$
- $\pi_t^k = \mathsf{softmax}\left[g_{\theta_k}\left(\mathbf{h}_{t-1}^s\right)\right]$
- $p(\mathbf{z}_t) = \prod_{k=1}^{K} \left(\pi_t^k\right)^{z_t^k}$
- $p(\mathbf{x}_t^k \,|\, \mathbf{x}_{<t}^k) = \mathcal{N}(\boldsymbol{\mu}_t^k, \mathrm{diag}\left[(\boldsymbol{\sigma}_t^k)^2\right])$
- $\mathbf{h}_t^k = f_{\varphi_k}\left(\mathbf{x}_t, \mathbf{h}_{t-1}^k\right)$

$$p(\boldsymbol{z}_{\leq T}, \boldsymbol{x}_{\leq T}) = \prod_{k=1}^{K} \prod_{t=1}^{T} \left(\pi_t^k \, p\left(\mathbf{x}_t^k \middle| \mathbf{x}_{<t}^k\right)\right)^{z_t^k}.$$

Neural Variational Switching Dynamical Systems

Inference



- $q(\mathbf{z}_t|\mathbf{x}_{\leq t}) = \prod_{k=1}^{M} \left(\rho_t^k\right)^{z_t^k}$
- $\mathbf{u}_k = \sigma\left(\mathbf{W}_k \, \mathbf{H}_t + \mathbf{V}_k \, \mathbf{h}_t^s + \mathbf{b}_k\right)$
- $\mathbf{H}_t = \left([\hat{\mathbf{x}}_t^1, \mathbf{h}_t^1], [\hat{\mathbf{x}}_t^2, \mathbf{h}_t^2], \ldots, [\hat{\mathbf{x}}_t^K, \mathbf{h}_t^K]\right)$
- $\rho_k^t = \mathsf{softmax}\left[\mathbf{u}_k \cdot \mathbf{c}_k\right]$

ELBO

$$\mathcal{L}[q] = \mathbb{E}_{p_D(\mathbf{x})}\left[\sum_{k=1}^{K}\sum_{t=1}^{T}\left\{\rho_t^k \log p(\mathbf{x}_t^k|\mathbf{x}_{<t}^k) + \rho_t^k \log\left[\frac{\pi_t^k}{\rho_t^k}\right]\right\}\right], \tag{8}$$

Mode regularization

$$\mathcal{H}[\rho] = -\mathbb{E}_{p_D(\mathbf{x})} \sum_{k=1}^{K} \tilde{\rho}_k \log \tilde{\rho}_k, \tag{9}$$

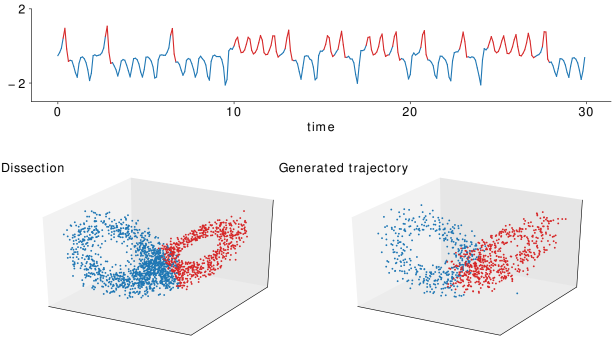
$$\mathcal{L}'[q] = \mathcal{L}[q] + \lambda_e \mathcal{H}[\rho] \tag{10}$$

An expectation-maximization solution to NVSDS (NVSDS-EM)

One can directly optimize by noticing $\mathcal{L}[q]$ is nothing but the negative Kullback-Leibler divergence between $q(\mathbf{z})$ and $p(\mathbf{x},\mathbf{z})$. An optimal lower bound is then found by minimizing this divergence. Such a minimum happens only for $\log q(\mathbf{z}) = \log p(\mathbf{x},\mathbf{z}) + \text{const}$.

$$q(\mathbf{z}) = \prod_{t=1}^{T} \prod_{k=1}^{K} \left(\frac{\rho_t^k}{\sum_k^K \rho_t^k} \right)^{z_t^k}, \quad \rho_t^k = \pi_t^k \, p(\mathbf{x}_t^k | \mathbf{x}_{<t}^k), \tag{11}$$

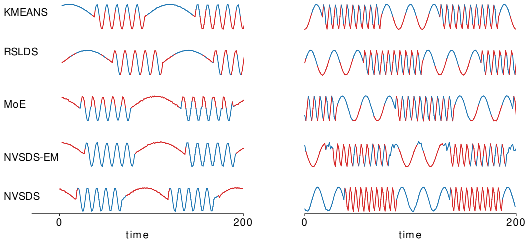
Lorenz Attracttor



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

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \sigma(y-x), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= x(\rho-z)-y, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = xy-\beta z. \end{aligned} \tag{12}$$

Switching Oscillatory Dynamics



It is defined as:

$$f(t) = H(\cos(\omega_s t) < 0)\cos(\omega_1 t) + [1 - H(\cos(\omega_s t) < 0)]\cos(\omega_2 t) \qquad (13)$$

Where H is the Heaviside function and $\omega_s < \omega_2 < \omega_1$

Handwritting

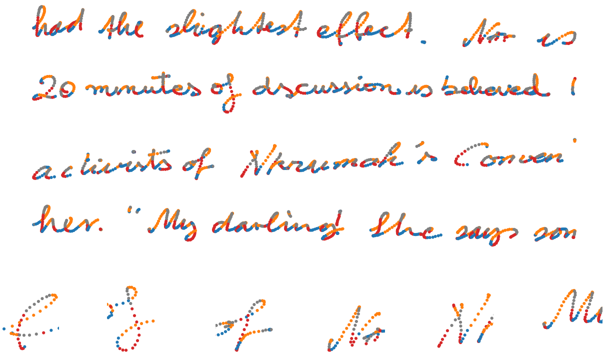


Figure: Dissection of a handwriting signal for the NVSDS model for different sequences. The lower row shows particular letters from the complete sequences for easier comparison.

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.