Flow-based Recurrent Mixture Density Network (FRMDN)

Ali Izadi

Table of Contents

Abstract

01 | Recurrent Mixture Density Networks (RMDNs) two main parts:

- Recurrent Neural Network (RNN)
- Gaussian Mixture Model

kind of RNN (almost LSTM) is used to find the parameters of a GMM

02 | estimating the covariance matrix for the high-dimensional problems is more difficult:

Tied configuration: $(\Sigma_k^{-1} = UD_kU)$ where U be an Identity and Dk is a specific diagonal matrix for kth component

03 | At every time step, the next observation, yt+1, has been passed through a flow-based neural network to obtain a much simpler distribution. Finally, diagonal GMM applied on transformed observations.

Introduction

01 | Generative Models(Flow based advantages):

- It can calculate the likelihood of data, while VAE only can approximate the lower-bound on the likelihood of data
- efficient training and inference methods
- they have access to latent space in comparison with GANs

02 | Mixture Density Network (MDN):

- parameters of GMM are generated by a neural network.
- its criterion is Negative Log-Likelihood (NLL)
- output of a neural network is coefficients, covariance matrices and mean vectors of the components in the mixture model.

03 | Recurrent Mixture Density Network (RMDN):

- RMDN uses a recurrent neural network to generate the parameters of a GMM in every time step to calculate the conditional probability of current observation given the history of inputs up to now.

Introduction

04 | Experimental results (Deep Reinforcement Learning)

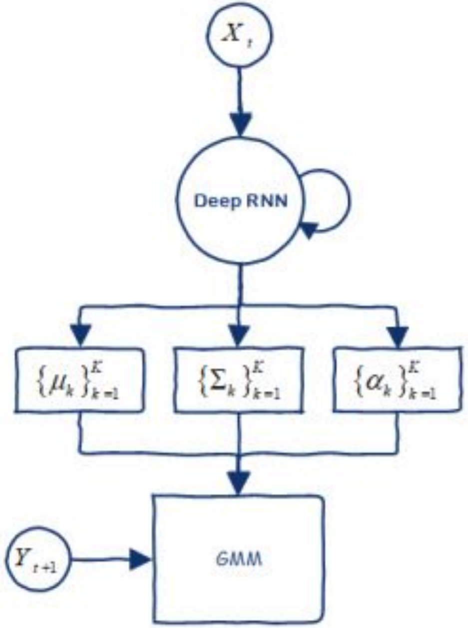
05 | Main contribution (MDN):

Since estimating a full covariance matrix is very complex due to hard optimization and overfitting problems. But, real-valued data is more complicated and RMDN needs a more powerful parametric distribution to model it.

- $(\Sigma_k^{-1} = UD_kU)$ every component in GMM has a specific diagonal matrix (Dk). Also there is a shared matrix (U) among components. So, there are different choices for the matrix U, based on its shape.

- We transfer original data through a flow-based neural network to gain much simpler distribution. Therefore, a simple form of RMDN (eg. RMDN with a diagonal GMM) for transformed data can be overcome this problem

Background (Recurrent Mixture Density Network)



Recurrent Mixture Density Network (RMDN)

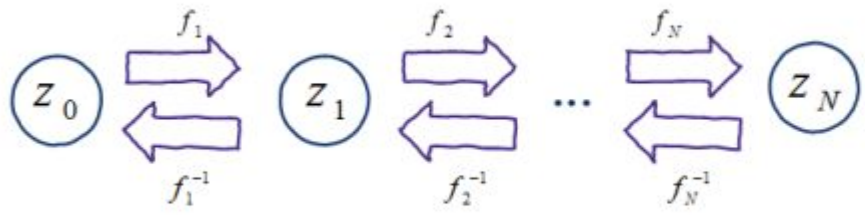
$$p(y_{t+1}|x_{\leq t}) = \sum_{k=1}^{K} \alpha_k(x_{\leq t})\varphi\left(y_{t+1};\theta_k(x_{\leq t})\right) \tag{1}$$

$$\varphi(y_{t+1};\theta_k(x_{\leq t})) = \frac{1}{(2\pi)^{d/2}|\Sigma_k(x_{\leq t})|^{1/2}} \exp\left\{\frac{-1}{2}(y_{t+1}-\mu(x_{\leq t}))^T\Sigma^{-1}(x_{\leq t})\left(y_{t+1}-\mu(x_{\leq t})\right)\right\} \tag{2}$$

$$NLL(Y|X) = -\sum_{q=1}^{Q}\log\prod_{t=1}^{T_q} p(y_{t+1}^q|x_{\leq t}^q) \tag{3}$$

there is a set of three kinds of parameters, mean vector (m), covariance matrix (S) and coefficient (a) of each component. Since these parameters must satisfy some constraints to be valid parameters, thus the last output of the neural network at each time step is divided into three parts and each part has its own specific activation function to satisfy these constraints.

Background (Normalizing flow)



Normalizing flow

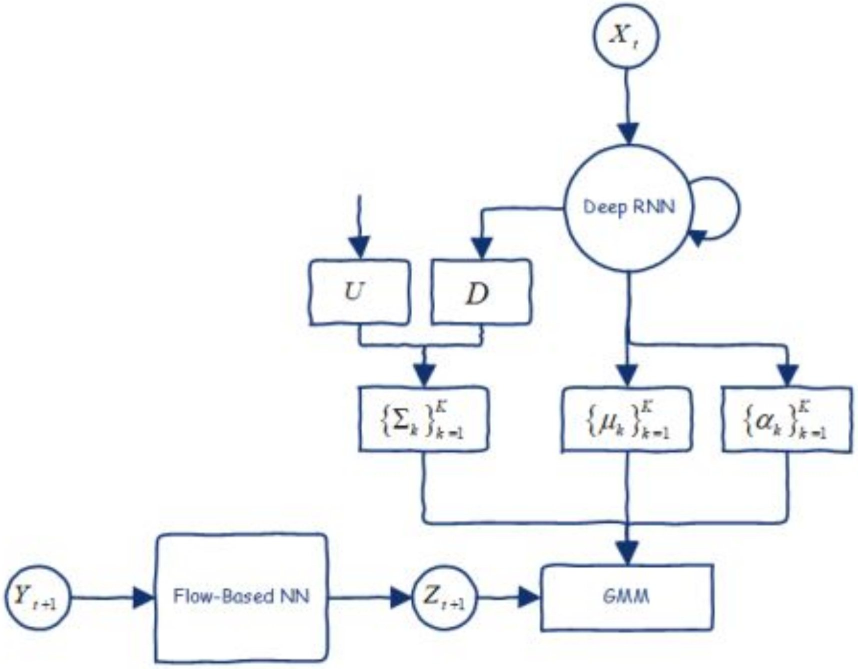
The main goal of NF is transferring a simple distribution to a more complex one.
It is done by applying a combination of **invertible functions.**

Proposed Method

$$f(y_{t+1}|x_{\le t}) = \sum_{k=1}^{K} \alpha_k(x_{\le t})\varphi(f(y_{t+1});\theta_k(x_{\le t})) \tag{9}$$

$$\varphi\left(f(y_{t+1});\theta_k(x_{\le t})\right) = \frac{1}{(2\pi)^{d/2}|\Sigma_k(x_{\le t})|^{1/2}} \exp\left\{\frac{-1}{2}(f(y_{t+1}) - \mu(x_{\le t}))^T UD(x_{\le t})U^T\left(f(y_{t+1}) - \mu(x_{\le t})\right)\right\} \tag{10}$$

$$NLL(Y|X) = -\sum_{q=1}^{Q} \log \prod_{t=1}^{T_q} \sum_{k=1}^{K} \frac{\alpha_k(x^q_{\le t})\,|U|\,\left|D_k(x^q_{\le t})\right|}{(2\pi)^{d/2}} \exp\left(\frac{-1}{2}\left(f(y^q_{t+1}) - \mu(x^q_{\le t})\right)^T UD(x^q_{\le t})U^T\left(f(y^q_{t+1}) - \mu(x^q_{\le t})\right)\right) \tag{11}$$



Flow based Recurrent Mixture Density Network

- U is an Identity matrix
- U is a square matrix
- U is a non-square matrix (future works)

References

- Razavi, S.F. and Hosseini, R., 2020. FRMDN: Flow-based Recurrent Mixture Density Network. *arXiv preprint arXiv:2008.02144*.

Thank you.