Hierarchical Deep Generative Models for Multi-Rate Multivariate Time Series

Zhengping Che

University of Southern California

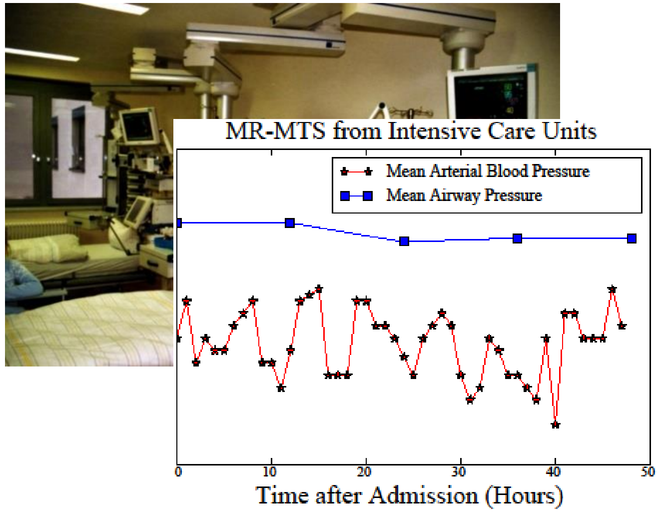
Joint work with Sanjay Purushotham, Guangyu Li, Bo Jiang, and Yan Liu

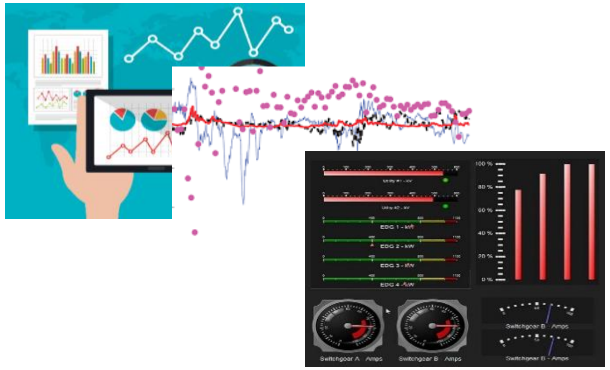
Jul 13, 2018

Poster **#54**
18:15 – 21:00

Introduction

- Multivariate Time Series (MTS) -- many real-world applications
  - Healthcare, climate, business analysis, financial forecasting, engineering…





- One of the key challenges -- Multi-Rate Multivariate Time Series (MR-MTS)
  - Different sampling rates
  - Multiple data sources / sensors

Motivation

- Major challenges of modeling MR-MTS
  - Need to handle *different* sampling rates
  - *Multi-scale* temporal dependencies
  - Complex underlying *generation* mechanism

- Existing solutions to MR-MTS forecasting/interpolation problems
  - *Single-rate* model? *(Kalman filter, VAR, deep Markov models, ...)*
    - Ignoring dependencies across different rates
  - Simple *imputations*? *(mean-imputation, Spline, MICE, MissForest, ...)*
    - May introduce unrelated/hide necessary dependencies
  - Multi-rate *discriminative* models? *(PLSTM, HM-RNN, ...)*
    - Not able to learn how the data is generated

Motivation

- Major challenges of modeling MR-MTS
  - Need to handle *different* sampling rates
  - *Multi-scale* temporal dependencies
  - Complex underlying *generation* mechanism

- Key point
  - To learn the **latent hierarchical structures** of the **data generation mechanism**

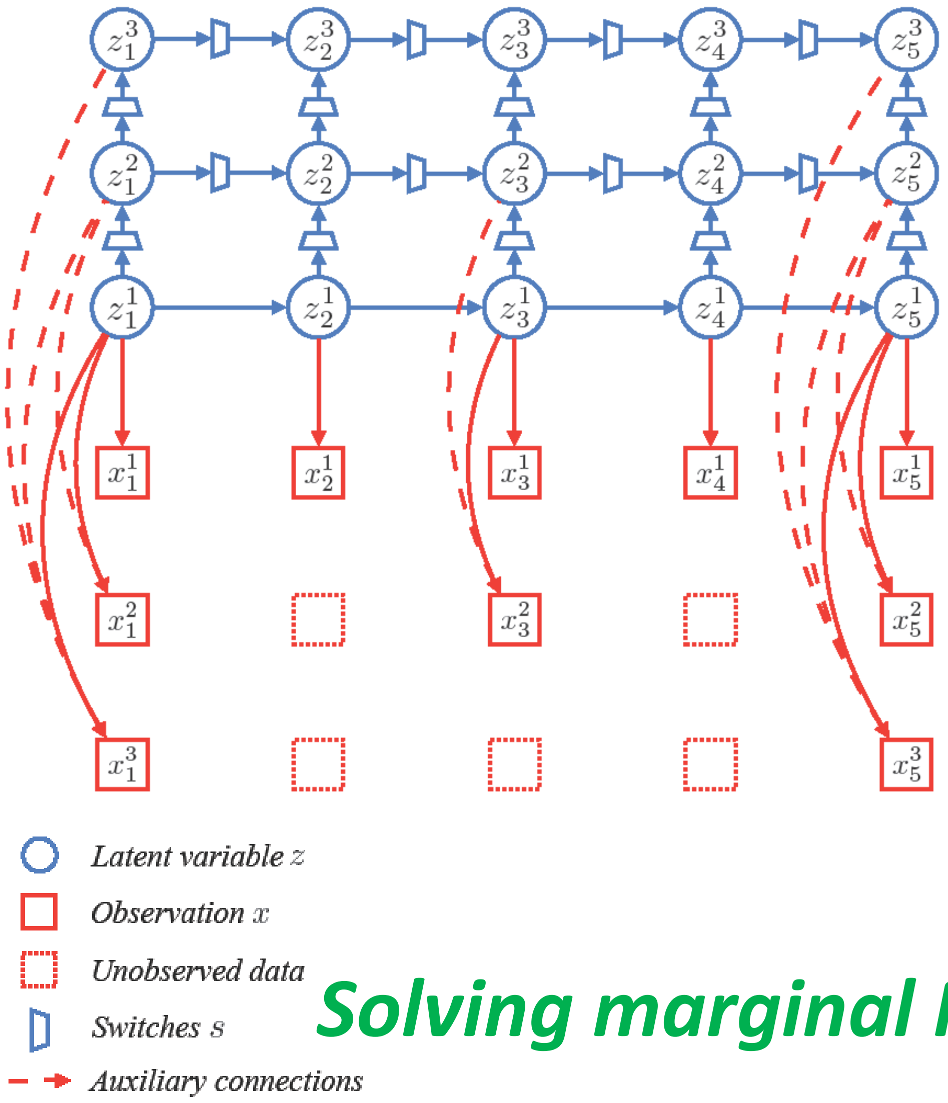
- Our proposed solution
  - **MR-HDMM**: **M**ulti-**R**ate **H**ierarchical **D**eep **M**arkov **M**odel

Overview

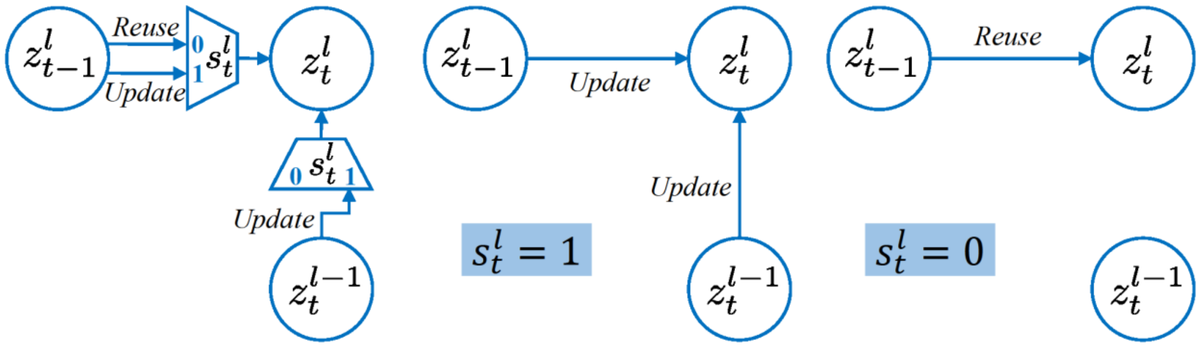
- Problem definitions
  - Input -- MR-MTS of L different sampling rates and T time steps ($\boldsymbol{x}_{1:T}^{1:L}$)
  - Case 1 -- forecasting problem
    - Output -- Given $x_{1:T}^{1:L}$, predict $x_{T:T'}^{1:L}$
  - Case 2 -- interpolation problem
    - Output -- fill-in missing values of lower sampling rates in $x_{1:T}^{1:L}$

- **MR-HDMM**: Multi-Rate Hierarchical Deep Markov Model
  - Component -- a **generation model** and an **inference network**
  - Motivation -- capturing hierarchical structures in underlying data generation process
    - **Learnable switches**
    - **Auxiliary connections**

Generation Model



- **Transition**
  - Learning latent states z
  - To capture **hierarchical structure**
    - Learnable switches
      - Update-and-reuse

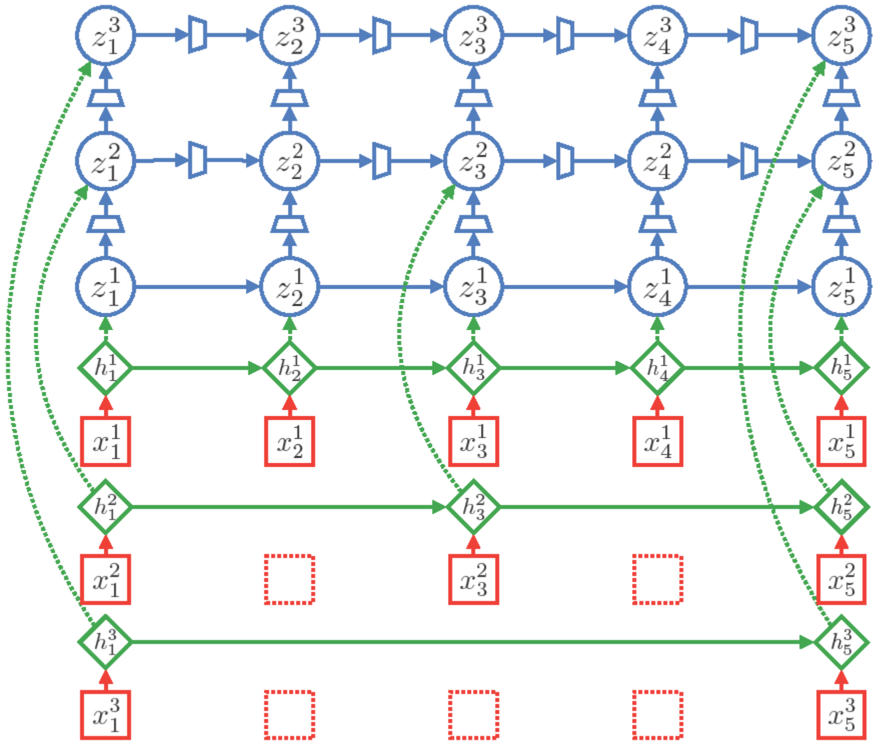


- **Emission**
  - Generating MR-MTS x
  - To capture **multi-scale dependencies**
    - Auxiliary connections

Solving marginal MLE?

Inference Network

- Keep similar structure as the generative model
  - Keeping the Markov properties of z
  - Inheriting the same switches s
  - Capturing MR-MTS observation by multiple RNNs
- Maximize the variational evidence lower bound (ELBO)
  - Conditional likelihood $\sum_{t=1}^{T}\sum_{l=1}^{L} \mathbb{E}_{Q^*(z_t^{1:l})} log\, p_{\theta_x}\left(x_t^l \,|z_t^{1:l}\right)$
  - KL at each time step and for each layer



$$\sum_{t=1}^{T}\mathbb{E}_{Q^*(z_{t-1}^1)} D_{KL}\left(q_\phi(z_t^1|x_{1:T}^{1:L}, z_{t-1}^1)||p_\theta(z_t^1|z_{t-1}^1)\right) + \sum_{t=1}^{T}\sum_{l=2}^{L} \mathbb{E}_{Q^*(z_{t-1}^l, z_t^{l-1})} D_{KL}\left(q_\phi(z_t^l|x_{1:T}^{1:L}, z_{t-1}^l, z_t^{l-1})||p_\theta(z_t^l|z_{t-1}^l, z_t^{l-1})\right)$$

Jointly learning all parameters

by stochastic backpropagation and ancestral sampling

Experimental settings

- Datasets

Domain	Dataset	# of Samples	Sampling Rates	# of Variables	Time Series Length
Healthcare	**MIMIC-III**	10709 (admissions)	1 / 4 / 12 Hours	7 / 12 / 44	72 Hours
Climate	**USHCN**	100 (years)	1 / 5 / 10 Days	70 / 69 / 69	365 Days

  - MIMIC-III: 5 runs × 5-fold CV (*train/valid/test split*)
  - USHCN: 5 runs of train/valid/test split with 1-month stride

- Forecasting baselines
  - **Single-rate**: Kalman Filter, VAR, Deep Markov Model, HM-RNN, LSTM, and PLSTM
  - **Multi-rate**: Multiple KF, Multi-Rate KF, and two simplified models of MR-HDMM
- Interpolation baselines
  - **Imputation**: Mean, CubicSpline, MICE, MissForest, SoftImpute
  - **Deep learning**: Deep Markov Model and the two simplified models of MR-HDMM

Quantitative results

- Forecasting

Method \ Dataset		MIMIC-III All	MIMIC-III HSR	MIMIC-III MSR	MIMIC-III LSR	USHCN All	USHCN HSR	USHCN MSR	USHCN LSR
Single-Rate Baselines	**Kalman Filter** (KF)	1.91×10^{18}	3.34×10^{18}	8.38×10^{9}	1.22×10^{6}	1.236	1.254	1.190	1.148
	Vector Autoregression (VAR)	1.233	1.735	0.779	0.802	2.415	2.579	1.921	1.748
	Deep Markov Model (DMM)	1.530	1.875	1.064	1.070	0.795	0.608	0.903	0.877
	HM-RNN	1.388	1.846	0.904	**0.713**	0.692	0.594	1.151	**0.775**
	LSTM	1.512	1.876	1.006	1.036	0.849	0.688	0.934	0.928
	PLSTM	1.244	1.392	1.030	1.056	0.813	0.710	0.870	0.915
Multi-Rate Baselines	**Multiple KF**	2.05×10^{18}	3.58×10^{18}	3.63×10^{4}	9.54×10^{2}	1.212	1.082	1.727	1.518
	Multi-Rate KF	1.691	2.289	0.944	0.860	0.628	0.542	0.986	0.799
	Multi-Rate DMM (MR-DMM)	1.061	1.192	0.723	1.065	0.667	0.611	0.847	0.875
	Hierarchical DMM (HDMM)	1.047	1.168	0.702	1.076	0.626	0.568	0.815	0.836
MR-HDMM		**0.996**	**1.148**	**0.678**	0.911	**0.591**	**0.541**	**0.742**	0.795

- Interpolation

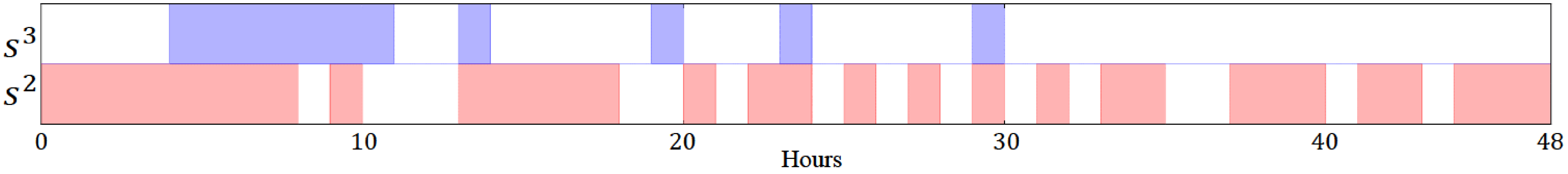
Method \ Dataset		MIMIC-III In-Sample	MIMIC-III Out-Sample	USHCN In-Sample
Imputation Baselines	**Simple-Mean**	3.812	3.123	0.987
	CubicSpline	3.713	3.212×10^{4}	0.947
	MICE	3.747	7.580×10^{2}	0.670
	MissForest	3.863	3.027	0.941
	SoftImpute	3.715	3.086	0.759
Deep Learning Baselines	**DMM**	3.714	3.027	0.782
	MR-DMM	3.710	3.021	0.696
	HDMM	3.790	3.100	0.750
MR-HDMM		**3.582**	**2.921**	**0.626**

HSR/MSR/LSR:
High/Mid/Low sampling rate

In/Out-Sample:
Interpolating training/testing dataset

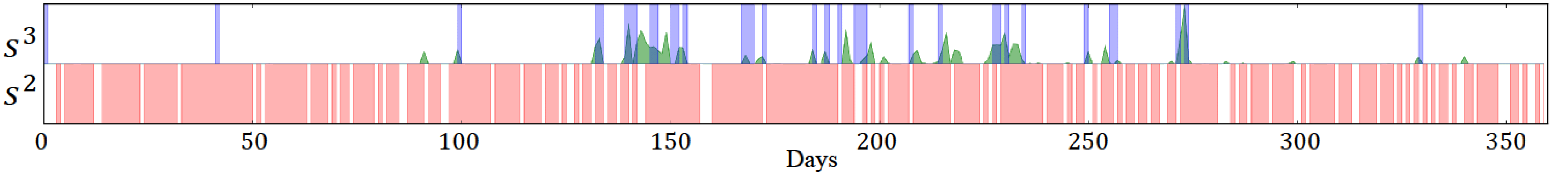
Visualizations of the learned latent hierarchical structures

- First 48 hours of an admission from MIMIC-III dataset



  - Blue: update of higher-layer states (s^3)
  - Red: update of lower-layer states (s^2)
  - Higher layer ⇒ fewer updates ⇒ longer-term dependencies

- A 1-year climate observation from USHCN dataset



  - Green: precipitation records
  - Changes in precipitations ⇒ significant differences ⇒ captured by the higher layer

Summary

- **MR-HDMM**: **M**ulti-**R**ate **H**ierarchical **D**eep **M**arkov **M**odel
  - A hierarchical deep generative model for multi-rate MTS
  - **Learnable switches** and **auxiliary connections** to learn the **latent hierarchical structures**
  - Variational inference methods to jointly learn all parameters
  - Excellent performance and visualizations on real-world datasets

- Future Work
  - To model MR-MTS with asynchronous/irregular sampling rates
  - To incorporate/justify domain-specific knowledge on data generation process
  - To build efficient model for large-scale MR-MTS datasets

Thank you!

Our poster is @ Hall B **#54**

Zhengping Che zche@usc.edu
USC Melady Lab http://www-bcf.usc.edu/~liu32/melady.html