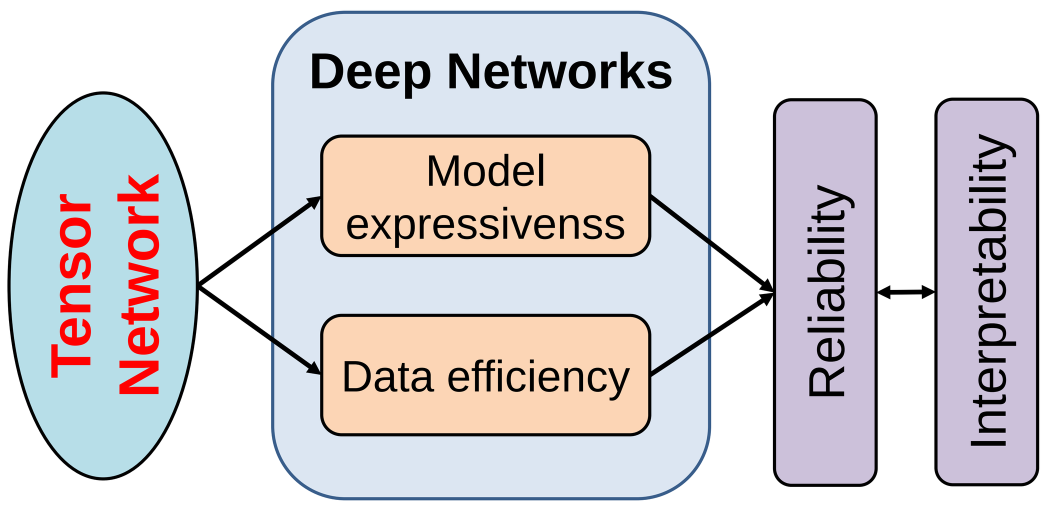
Tensor Learning Team
Qibin Zhao
テンソル学習チーム　趙 啓斌

RIKEN

AIP Center for Advanced Intelligence Project

Research Outline

- Background: Modern ML requires a huge amount of data and large model sizes, which is also vulnerable to adversarial attacks and lack of interpretability.
- Goal: To study innovative machine learning model and algorithms by using tensor representation, tensor decomposition/networks with aim to address the challenges of data efficiency, model expressiveness with efficient parameters, robustness to adversarial attacks, and interpretability.

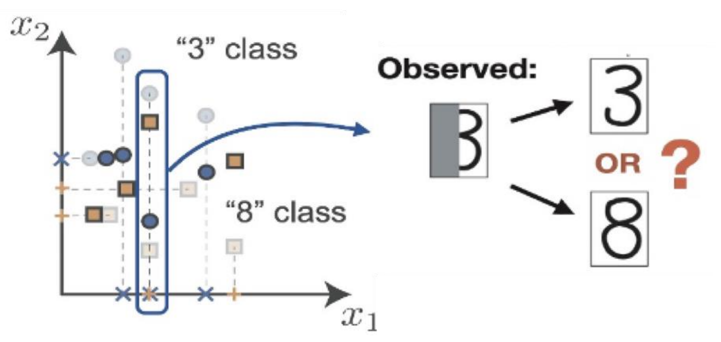


Learning with Incomplete or Limited data

- Learning representation and classification model from incomplete data (Caiafa et al., CVPR workshop 2021)

Objective: $\hat{f}(g(x^{miss}), \hat{\theta}) \approx f(x, \theta)$

Reconstruction of incomplete data



Approach: Simultaneous reconstruction (sparse coding) and classification (DNNs) with sufficient condition

$$J(\Theta, \mathbf{D}, \mathbf{s}_i) = \frac{1}{I}\sum_{i=1}^{I}\{J_0(\Theta, \hat{\mathbf{x}}_i, y_i) + \lambda_1 J_1(\mathbf{D}, \mathbf{s}_i) + \lambda_2 J_2(\mathbf{s}_i)\}$$

Classification loss (e.g. crossentropy) for any classifier (deep network)

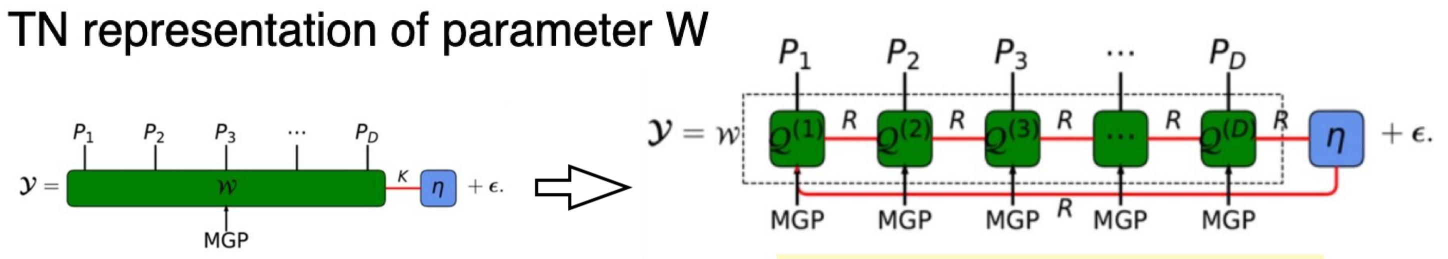
Representation error $J_1(\mathbf{D}, \hat{\mathbf{s}}_i) = \frac{M}{N}\|\mathbf{m}_i * (\mathbf{x}_i - \mathbf{D}\mathbf{s}_i)\|^2$.

Promotes sparsity $J_2(\mathbf{s}_i) = \frac{1}{N}\|\mathbf{s}_i\|_1$

- Latent factor analysis with limited data samples (Tao et al., ACML 2021)

$$\boldsymbol{y}^{(n)} = \boldsymbol{W}\boldsymbol{\eta}^{(n)} + \boldsymbol{\epsilon}^{(n)}, \quad \forall n = 1, \ldots, N,$$

- Given higher-order data $\mathcal{Y} \in \mathbb{R}^{P_1 \times \cdots \times P_D}$, marginalize η gives $\mathcal{Y} \sim \mathcal{N}(0, \mathcal{V})$
- TN representation of parameter W

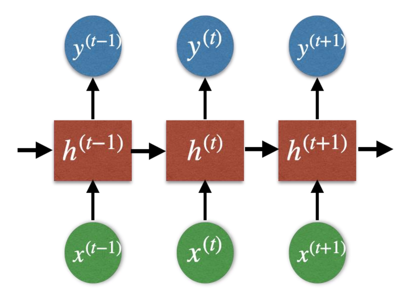


$$\mathcal{Y} = \ll \mathcal{Q}^{(1)}, \ldots, \mathcal{Q}^{(D)}, \boldsymbol{\eta} \gg + \boldsymbol{\mathcal{E}},$$

- Symmetric tensor ring reparameterization
- Capture structural dependency within data

Expressiveness with Tensor Networks

- RNN and LSTM do not have long memory from a statistical perspective [Zhao et al., ICML 2020]
- Challenge: How to achieve long memory?



$h^{(t)} = \sigma(Wh^{(t-1)} + Ux^{(t)} + b)$

Tensor-Power Recurrent Models (Li et al., AISTATS 2021)

$(p+1)$-order weight tensor

$$\mathbf{h}^{(t)} = \mathcal{G} \times_1 \begin{pmatrix}\mathbf{x}^{(t)} \\ \mathbf{h}^{(t-1)}\end{pmatrix} \times_2 \cdots \times_p \begin{pmatrix}\mathbf{x}^{(t)} \\ \mathbf{h}^{(t-1)}\end{pmatrix} = \mathcal{G} \cdot \underbrace{\begin{pmatrix}\mathbf{x}^{(t)} \\ \mathbf{h}^{(t-1)}\end{pmatrix}^{\otimes p}}_{p\text{-fold tensor product with itself}}$$

Large p – long memory, high expressiveness

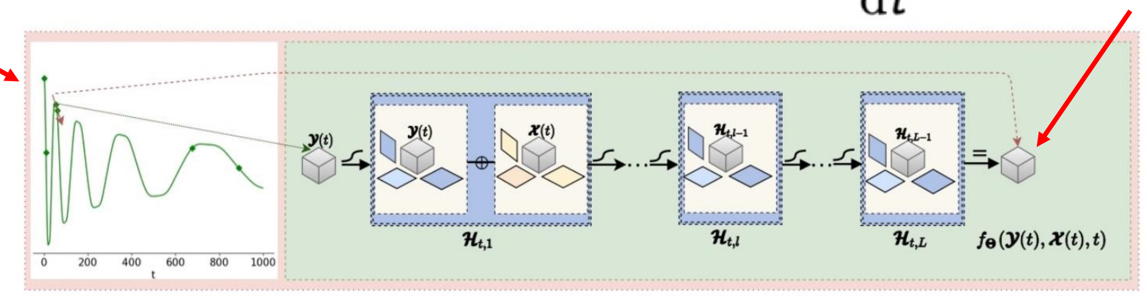
STD factors of weight tensor

$$\mathbf{h}^{(t)}[j] = \sum_{r=1}^{R}\left\langle \mathbf{w}_{j,r}, \begin{pmatrix}\mathbf{x}^{(t)} \\ \mathbf{h}^{(t-1)}\end{pmatrix}\right\rangle^p + \mathbf{b}[j],$$

- Given tensorial time series with irregular time steps, how to achieve prediction on continuous time points with efficient parameters?

Tensor Neural ODE (Bai et al., IJCNN 2021)

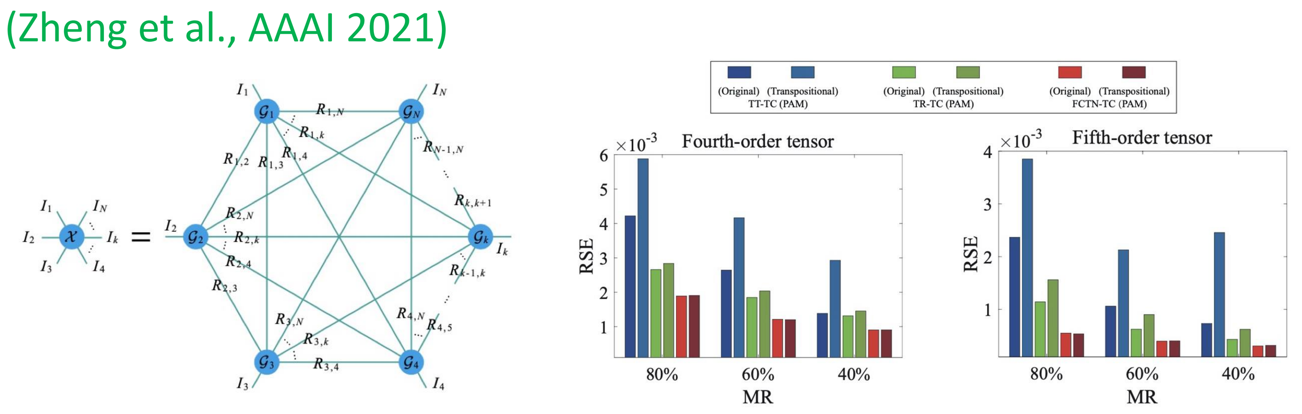
$$\{\mathcal{Y}_t\}_{t\in[0,T]}, \mathcal{Y}_t \in \mathbb{R}^{I_1 \times \cdots \times I_N} \qquad \frac{\mathrm{d}\mathcal{Y}(t)}{\mathrm{d}t} = f_{\Theta}(\mathcal{Y}(t), \mathcal{X}(t), t)$$



TENODE: architecture with tensor contraction layer

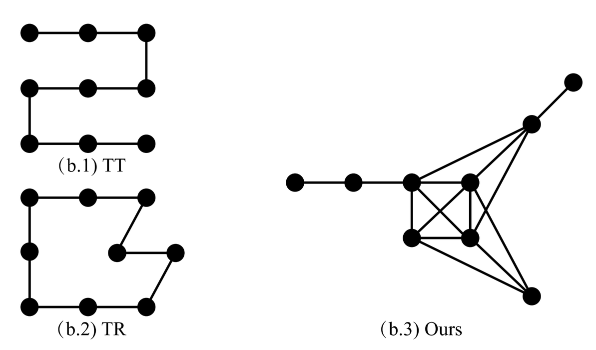
Tensor Network Topology Learning

- Fully Connected TN (FCTN) for tensor completion (Zheng et al., AAAI 2021)



- Fully capture correlations between pairwise modes
- Transpositional Invariance

- Problem: The existing TN structures are *NOT* the optimal, and the searching task is *NP-hard*.



known structures

More efficient structures

$$\min_{\mathbf{A}\in\mathbb{A}} \frac{1}{\epsilon(\mathbf{A})}, \quad s.t.\ \exists\hat{\mathbb{V}} \text{ which satisfies } \left\|\mathcal{X} - TN(\hat{\mathbb{V}}; \mathbf{A})\right\|_F^2 \leq \delta.$$

Collection of core tensors

Adjacency Matrix

$$\begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

Order-4 tensor with TT-format

$$\epsilon(\mathbf{A}) = \frac{\text{Uncompressed size of } \mathcal{X}}{\text{Parameter size of } \mathbb{V} \text{ under } \mathbf{A}}.$$

(Li et al., ICML 2020)

Learn (near-)optimal TN topology via evolutionary algorithm (EA).

References

- H. Qiu, C. Li, Y. Weng, Z. Sun, X. He, and Q. Zhao. 2021, "On the memory mechanism of tensor-power recurrent models," AISTATS 2021.
- Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Q. Zhao, and T.-X. Jiang. "Fully-connected tensor network decomposition and its application to higher-order tensor completion," AAAI 2021.
- Z. Tao, X. Zhao, T. Tanaka, and Q. Zhao. "Bayesian latent factor model for higher-order data," ACML 2021.
- C. F. Caiafa, Z. Wang, J. Sole-Casals, and Q. Zhao. "Learning from incomplete features by simultaneous training of neural networks and sparse coding," CVPR Workshops 2021.
- M. Bai, Q. Zhao, and J. Gao. "Tensorial time series prediction via tensor neural ordinary differential equations," IJCNN 2021.
- Co-organizer of NeurIPS 2021 second workshop: "Quantum Tensor Networks in Machine Learning" https://tensorworkshop.github.io/NeurIPS2021/index.html.
- Our team's webpage: https://qibinzhao.github.io.