Collaborators: **Tsuyoshi Okubo** (Univ. of Tokyo), **Naoki Kawashima** (ISSP)

Tensor tree learns hidden relational structure in data to construct generative models

Nov. 16, 2024

Graduate School of Informatics, Kyoto University Kenji Harada

This study is supported by **TOYOTA**

Workshop: Tensor network 2024 (Kanazawa)

Applications of TN techniques to machine learning

I.V. Oseledets, "Tensor-Train Decomposition," (2011).

Tensor decomposition in machine learning algorithms

• Tensor train (= **MPS**: matrix product state)

- Compression of tensor in machine learning algorithms E. Stoudenmire, and D.J. Schwab, "Supervised Learning with Tensor Networks," (2016).
- Quantics tensor train

- Images: Latorre (2005), …
- NN: Novikov, et al. (2015), …

$$
x = x_N \times 2^{N-1} + x_{N-1} \times 2^{N-2} + \dots + x_2 \times 2 + x_1
$$

Tensorization
$$
x \to (x_N, x_{N-1}, \dots, x_1)
$$

The number of TN applications rapidly grows in machine learning!

Generative modeling is an important technique in machine learning

Generative modeling involves creating a classical distribution model behind data.

Models for generative modeling

Models for generative modeling

- **Boltzmann** machine and restricted Boltzmann machine (RBM)
- Variational autoencoder (VAE)
- Generative adversarial network (GAN)
- Normalizing flow
- Diffusion model

Diffusion process

Based on projective measurements of

a **quantum** state

$$
p(\mathbf{x}) = |\psi(\mathbf{x})|^2
$$

Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, Phys. Rev. X **8**, 031012 (2018).

• Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang, "Unsupervised Generative Modeling Using Matrix Product States," Physical Review X, **8**, 031012(2018). *hb*. The obtained matrices *A* and *B* can be absorbed into a ng Using Matrix Product States," Physical Review X, **8**, 031012(2018). $\frac{1}{3}$ bong matrix ribudut otates, i hydical Keview X, \bullet , colorz (2010) .

• Marcello Benedetti, *et al.*, "A generative modeling approach for benchmarking and training shallow quantum circuits," npj Quantum Information, **5**, 45(2019). literature and will be contracted later. The exposed edge denotes the raining shallow quantum circi be determined by the data set. For one of the given configurations of of noi Quentum Informatio As in the International control in the International control in the Turk generation in the Turk generation in the Tu quantum circuits," npj Quantum Information, **5**, 45(2019).

g Borr that is, the TTN represents a pure quantum state "(*x*), and λ information $\boldsymbol{6}$ 60(2020) $\mathcal{L}^{\mathcal{L}}$ adjust dynamically the bond dimensions of \mathbf{v} , \mathbf{v} ,

network has corresponding factor graphs [25]. We take the plied, **16**, 044057(2021).

M(*k*) in each edge *k* and an identity tensor δ(*j*) in each hidden via tensor networks," Nature Communica[.] etworks." Nature Communications, **14**, 8367(2023).

ገ*ንለ* ነ \ddot{H} \ddot{H} are \ddot{H} are \ddot{H} are key differences that there are key differences that there are key differences that the set of \ddot{H} are \ddot{H} are \ddot{H} are \ddot{H} are \ddot{H} are \ddot{H} are \blacksquare

For *j* = 1, we simply let the bond dimension of *z, u* equal 1. ang, T. Xiang, and Pan Zhang, "Tree tensor networks for generative modeling," Physical Review B, **99**, 155131(2019).

> \blacksquare $T = T$ ivir $R A$, $P P P D$, P \cdots in the sum-product algorithm. 1λ MERA, PEPS, ...

 η rigarding the edge between the edge between the edge between the edge between two tensors η $t \sim \tau$ hereafter. The left and right indices of the tensors in the Review A, **98**, 062324

Two approaches for the Born machine

Tensor network model (TN) dimension increases, a MPS enhances its ability to para-

- n Wang, Heng Fan, Lei Wang, and Pan Zhang, "Unsupervised Generative Modeling Using Matrix Product States,"
 $r_{\text{max}} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=$
- Song Cheng, Lei Wang, T. Xiang, and Pan Zhang, "Tree tensor networks for generative modeling," Physical Review B, 99, 155131(2019).

Parametrized quantum circuit model (PQC) \blacksquare a random variable with a value of 1 block represents a hidden (visible) variable. The edge between two \mathbf{r} ized quantum circuit model (POC) indicate tensor contraction over virtual indices. The dangling

- Jin-Guo Liu and Lei Wang, "Differentiable learning of quantum circuit Born machines," Physical Review A, **98**, 062324(2018). Lei Wang, "Differentiable learning of quantum circuit Born machines," Physical Review A, **98**, 062324(2018).
- R ful, et al., "A generative modelling approach for benchmarking and training.
.
- Brian Coyle, et al., "The Born supremacy: quantum advantage and training of an Ising Born machine," npj Quantum Information, 6, 60(2020). l., "The Born supremacy: quantum advantage and training of an Ising Born machine," npj Quantum Information, **6**,
- Marcello Benedetti, et al., "Variational Inference with a Quantum Computer," Physical Review Applied, **16**, 044057(2021). et al., "Variational Inference with a Quantum Computer," Physical Revie
- Manuel S Rudolph, *et al.*, "Synergistic pretraining of parametrized quantum circuits via tensor networks," Nature Communications, **14**, 8367(2023). which allows one to restrict the tensors with canonical $\mathcal{L}_{\mathcal{A}}$ on, et al., "Synergistic pretraining of parametrized quantum circuits via ten
- Mohamed Hibat-Allah, et al., "A framework for demonstrating practical quantum advantage," Communications Physics, **7**, 68(2024). -Allah, et al., "A framework for demonstrating practical quantum advantage," Communications Physics, **7**, 68(2024).

Synergistic pretraining of parametrized quantum circuits via tensor networks of the space head an mate poromotrized cuenture eiro tito u parances izuu quaritum virustud

Manuel S Rudolph, et al., Nature Communications, **14**, 8367(2023). for success using the best solution available with the best solution available with today's abundant today's a
The best solution available with the best solution and the best solution and the best solution and the best so might finally understanding unlock the true potential of parameters \mathbf{r}_i

Synergistic approach by TN and PQC \mathcal{L} unbounded), and criterion (c) needed to avoid fidelity field to avoid fidelity field f \blacksquare by nergistic approach by algorithms as effective methods for solving problems of deep problems of deep problems \mathcal{L}

Time on Classical Hardware Time on Quantum Hardware

resources (ref. 32 is limited here), criterion (b) needed to use the limited to use to use to use the use of u
In the criterion (b) needed to use the use of use to use the use of use to use the use of use the use of use t

 $\frac{1}{2}$ on $\frac{1}{2}$ dataset are high-set are high-set are high, the resulting $\frac{1}{2}$

The synergistic approach resolves the issue of the prevalence of barren plateaus in PQC **of the prediction dendscapes** exhibit significantly more stable behavior. After significant significantly more stable behavior. After stable behavior. After stable behavior. After significant significantly more stabilitie The gradient magnitude, i.e., the 2-norm of the 2-norm of the 2-norm of the gradient vector, is then evaluated on magnitude of the randomly initialized circuits decay exponentially with the number $\overline{\mathcal{P}}$, $\overline{\mathcal{P}}$, the gradients for the $\overline{\mathcal{P}}$ start to decay and are sur-are sur-are

dimensions $\mathcal{L} = 2$ or $\mathcal{L} = 4$ or $\mathcal{L} = 4$ or $\mathcal{L} = 4$, decompose them into one layer of $\mathcal{L} = 4$

Network structure and performance

$$
\mathcal{L} = -\frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \ln[p(\mathbf{x})] = -\frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \ln|\Psi(\mathbf{x})|^2
$$

TN of th or and quantum dealers $\frac{1}{2}$ **TN of the quantum state for the Borm machine**

The binary tree reaches the optimal value of NLL.

Negative Log-likelihood (NLL) = KL-divergence - entropy of data

S. Cheng, L. Wang, T. Xiang, and P. Zhang, Phys. Rev. B, **99**, 155131(2019). **Loss function**

K structure with good performance

Use of prior knowledge of data The line on the line on the right side represents the identity matrix. The identity matrix \mathcal{L}_max

 \det data, how can we design a good network structure? where the orange tensor represents the non-canonical central central central central central central central ce a daeign a good na sors' canonical forms is pointed toward the direction of the modeling of natural images. Figure 2(a) shows the two- $\mathsf{M}\cap\mathsf{r}\mathsf{V}$ etrional modeling $\mathsf{M}\cap\mathsf{r}\mathsf{V}$ $\mathbf v$ is $\mathbf v$ is responsible for $\mathbf v$ and $\mathbf v$ are $\mathbf v$ is a set \math

.

It is technically easy to canonicalize a tensor in the TTN.

Optimizing a tensor network structure

- In the case of the ground-state calculation, the network structure
	- $Ex. 1D \text{ model} \Rightarrow \text{MPS}, 2D \text{ model} \Rightarrow \text{PEPS}$
		-

aligns with local interactions on a lattice.

Usually, we first fix a tensor network structure.

However, we often have **no prior** knowledge of data.

Our goal

Optimizing a tensor network structure for generative modeling without prior knowledge of data.

Optimization of network structure for a ground-state calculation

Select a new decomposition with **the least entanglement entropy Step III**

T. Hikihara, H. Ueda, K. Okunishi, **K.H.**, and T. Nishino, Phys. Rev. Research **5**, 013031 (2023)

Truncation is small

Based on a two-sites algorithm of DMRG

In the class of **general tree** TNs

• Visualization of entanglement structure

Results of optimization of network structure for a ground-state calculation

• Improvement of variational energies

T. Hikihara, H. Ueda, K. Okunishi, **K.H.**, and T. Nishino, Phys. Rev. Research **5**, 013031 (2023) **AUTOMATIC STRUCTURE 2023**, N. Oeda, K. Okamsin, **K.m.**, and T. Nishino, I hys. Kev. Kesearch **3**, 013031 (2023)

Inhomogeneous AFH

Classical mutual information

$$
I(A:B) = \sum_{(a,b)} P(a,b) \ln \left[\frac{P(a,b)}{P(a)P(b)} \right]
$$

Classical mutual information

$$
I(A:B) = \sum_{(a,b)} P(a,b) \ln \left[\frac{P(a,b)}{P(a)P(b)} \right].
$$

Entanglement entropy

 \leq E.E.

Classical mutual information

\leq E.E. \leq ln(*D*) in a tree

$$
I(A:B) = \sum_{(a,b)} P(a,b) \ln \left[\frac{P(a,b)}{P(a)P(b)} \right].
$$

Entanglement entropy

Classical mutual information

$$
I(A:B) = \sum_{(a,b)} P(a,b) \ln \left[\frac{P(a,b)}{P(a)P(b)} \right]
$$

Selecting a new decomposition with the least classical mutual information

Adaptive tensor tree generative modeling

Select a new decomposition with **the least classical mutual information**

K.H., Tsuyoshi Okubo and Naoki Kawashima, arXiv:2408.10669

In the class of **general tree** TNs

Stochastic estimation of mutual information

Probabilities for a distribution and **marginal** distributions can be calculated for a tree TN.

Four applications of the ATT method

-
- Images of hand-written digits (QMNIST) • Ten random bit sequences with long-range correlations
- Bayesian networks
- Stock-price fluctuations in the S&P 500 index

Tensor tree learns hidden relational structure in data to construct generative models

Application to random bit sequences with long-correlations

Application to images of digits

Data with a random permutation of pixels

Results of NLL

Optimized network structure

The ATT method automatically learns the relevant relational structure among the random variables and places them close together on the tensor network

Application to Bayesian network's data

Graphical model (Bayesian network)

$$
P_{\text{data}}(\mathbf{x}) = \prod_{i} P(x_i | \{x_p\}_{p \in \text{Parent}})
$$

Causal dependencies among random variables

The ATT method successfully captures the corresponding topology of Bayesian networks

Application to stock-price fluctuation in S&P 500 index

NLL in training process

Data: binarized change rates of stock prices 1 if it is higher than the average for all stocks and 0 otherwise.

NLL vs. bond dimension

Optimized networks achieve better performance.

Optimized network structure for stock-price fluctuation in S&P500 index

Optimized network structure for stock-price fluctuation in S&P500 index

- Adapted tensor tree generative modeling
	- Succeeds for data with no prior knowledge
	- Optimized network structure shows hidden relational structure behind data

Tensor tree learns hidden relational structure in data to construct generative models

Reference: **arXiv:2408.10669**

Sample code:

https://github.com/KenjiHarada/adaptive-tensor-tree-generative-modeling

