Neural Network and Tensor Network Quantum States

Author Lu Wei

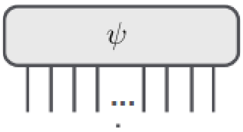
University USTC

November 3, 2021

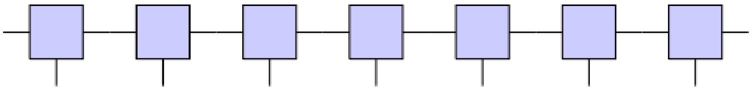
Contents

1 Tensor Network and Entanglement Area Law

Here we focus on the representation power of TN and MPS. Usually the quantum state of a multi-body system with n particles is an n-order tensor: Set every particle's local Hilbert space to have a dimension of d would be



a big tensor to have in total d^n parameters. It's impossible to deal with if n is large. But we can disassemble it into multiple small tensors. MPS is one kind of TN. In MPS: Each particle takes over one position. Interaction



between two adjacent particles can be represented by bond contraction. Use x to represent the particle and A every matrix, then the quantum state can be written:

$$|\Psi\rangle = \sum_{x_1,\cdots,x_n} c_{x_1,x_2,\cdots,x_n} |x_1, x_2, \cdots, x_n\rangle$$

$$c_{x_1,x_2,\cdots,x_n} = \sum_{a_1,\cdots,a_{n-1}} A^{x_1}_{a_1} A^{x_2}_{a_1,a_2} \cdots A^{x_{n-1}}_{a_{n-2},a_{n-1}} A^{x_n}_{a_{n-1}}$$

Denote connection dimension between two tensors by D, physical dimension by d, then the number of all parameters in n tensors is ndD^2 which is much smaller than the number of the whole parameters d^n. We know that the space of the latter one is much smaller than the space of the first one. We see the MPS is naturally suitable for one-dimensional system. If we want to apply it to two-dimensional system, we have to artificially appoint a one-dimensional chain of the physical system to bunch the whole particles in this system. However, this way will cause some adjacent particles to get far away in MPS thus adjacent interation would be far away also. It's obvious that the effect would be quite bad.

Then we turn to entanglement entropy for the quantitative explanation. For a quantum system conposed of A and B partition, its wave function is $|\Psi_{AB}\rangle$, and the corresponding subsystem A and B 's reduced density matrix are respectively ρ_A and ρ_B. Entanglement entropy's definition is:

$$S_{A|B} = -\operatorname{Tr} \rho_A \log_2 \rho_A = -\sum_i \lambda_i \log_2 \lambda_i \tag{1.1}$$

where λ_i is eigenvalue of ρ_A. The definition is related to schmidt of quantum state: suppose $|\Psi_{AB}\rangle$ has schmidt decomposition:

$$|\Psi_{AB}\rangle = \sum_i \alpha_i |A_i\rangle \otimes |B_i\rangle \tag{1.2}$$

where $\{|A_i\rangle\}$ and $\{|B_i\rangle\}$ are respectively a set of standard orthogonal bases of subsystem A and B and α_i are schmidt coefficients. We write it in the form of density matrix and compare with equation 1.1 and find that ρ_A and ρ_B have the same eigenvalue spectrum: $\lambda_i = |\alpha_i|^2$. So equation 1.1 holds for A and B with the same value.

If $|\Psi_{AB}\rangle$ is separable which is $|\Psi_{AB}\rangle = |A\rangle \otimes |B\rangle$, then there is no entanglement between A and B corresponding to 1.2 having only the first item.and $|\Psi_{AB}\rangle = |A\rangle \otimes |B\rangle$. When there is quantum entanglement between A and B, 1.2 would be summation and the corresponding entanglement entropy is the shannon entropy of this probability distribution. When $|\alpha_1|^2 = |\alpha_2|^2 = \cdots = |\alpha_n|^2 = 1/n$, the entenglement entropy of the system reach its maximum $\log_2 n$.

So, entanglement entropy quantitatively evaluate the correlation between two systems in a quantum system. In real physical system, the correlation shows on the interaction between particles. In particular, if you devide the whole physical system into two parts, the entanglement between them should be related to the strength and way of interaction. We have entanglement area law:

> **Theorem 1.1.** For local interation and gapped quantum many-body system, the entanglement entropy between any two parts of it is proportional to the contact area between the two parts.

Most quantum many-body system satisfies entanglement area law:

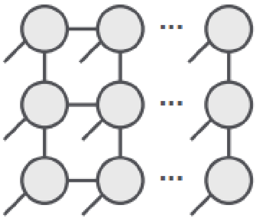
- One-dimensional chain: devide it into two parts and the contact place is always one point; However the extension of the chain, the entanglement entropy is always a constant: $S \propto O(1)$.
- Two-dimensional chain: devide it into two parts from the center and the contact between them will be a line; set the side two have length L and the contact line's length would have a length of $O(L)$. Thus the entanglement entropy would increase with the growth of the system: $S \propto O(L)$.

To calculate the entenglement entropy of MPS, we have to devide it into two parts and do schmidt decomposition on it. The process can be called MPS regularization and the algorithm touched on SVD. After that the left and right part would respectively become standard orthogonal bases and we get a set of singular values of the matrix in the cut point which is the schmidt coefficients in 1.2.

MPS has connection dimension D. So after decomposition we have at most D singular values. Calculate by 1.2 and we find that the maximum the entanglement entropy can reach is $\log_2 D$. So for one dimensional physical system with $S \propto O(1)$, D does not increase with the system size. Thus MPS can accurately calculate the quality of large one-dimensional physical system. However, for two dimensional physical system, $S \propto O(L)$ and correspondingly $D \propto 2^L$. That is to say, to fully describe the entanglement quality of a two dimensional physical system, the connection dimension would correspondingly increase expomentially with the length of the system which is way beyond the computation ability of computers. So MPS is suitable for one-dimensional system but is helpless for systems of dimension of two or above two.

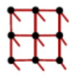
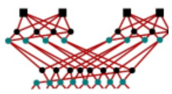
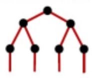
2 Representation Power

From previous analysis we see that MPS can only represent physical system of constant entanglement entropy. In order to calculate physical system of two or more dimension, some proposed the generalization of MPS in higher dimension which is PEPS (Projected entangled pair state):



Compared with MPS, PEPS's structure change from one-dimensional chain to two-dimensional plane. So small tensors from PEPS correspond to particles in two dimensional physical system.

Obviously the representation power of PEPS is stronger than that of MPS. Generally, for a new ansatz wave function, people cares about its representation power and it can be analyzed by its entanglement entropy. Some usual tensor network and the maximum entanglement entropy it can represent.

Tensor network structure	Entanglement entropy $S(\mathcal{A})$	correlation length ξ	local observable $\langle \hat{O} \rangle$	diagram
Matrix product state	$O(1)$	finite	exact	
Projective entangled pair state ($2d$)	$O(\lvert\partial\mathcal{A}\rvert)$	finite/infinite	approximate	
Multiscale entanglement renormalization ansatz ($1d$)	$O(\log\lvert\partial\mathcal{A}\rvert)$	finite/infinite	exact	
Branching multiscale entanglement renormalization ansatz ($1d$)	$O(\log\lvert\partial\mathcal{A}\rvert)$	finite/infinite	exact	
Tree tensor networks	$O(1)$	finite	exact	

Also entanglement entropy can evaluate the representation power of neural network as it's hard to analyze the theory of tensor network. The most important result is from reference 3. It analyzed the entanglement quality of RBM state and the result is RBM with only short-range connection satisfies entanglement area law while RBM with long-range connection satisfies entanglement volume law which means that the representation power of RBM exceeds almost all kinds of tensor network. What's more, RBM is only a two-layer neural network which means that the representation power of most other tensor networks is stronger than that of RBM.

Another important network structure is convolutional neural network. CNN's neurons of convolution layer and pooling layer have only spatially adjacent interaction. So it's similar to RBM with short-range RBM and satisfies entanglement area law. But there are many convolutional layers in convolutional neural network and every convolutional layer make the interaction range between neurons lager. Usually cnn can catch long-term correlation information in image identification task. In this case the combination of local interaction forms long-term correlation and a topic is that whether the maximal entanglement entropy it can encode can break area law.

Anyway, fully-connected layer means volume law which makes the representation power of tensor network exceeds all kinds of tensor network.

3 Entanglement Quality of Image Data Set

Opon previous discussion, there are some interesting issues:

- Entanglement qualities of Image Data Set
- What kind of ansatz wave functions can represent an Image data set effectively

Most real physical systems are local gap system and thus satisfies area law. However there is large amount of long-term correlation in images and it's far away from local interaction in real physical system. For example, figure images are all symmetric so we can speculate by half of the face the next half of the face. It can be done by local interactions so there must be long-term correlations which has always been a hard problem in physics and the complexity of image data set is much lager than that of real physical system.

Long-term interactions means entanglement volume law. Suppose we have a quantum state with two identical subsystem and the configuration of each subsystem is totally random. Such kind of quantum state can be written:

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_i |\psi_i\rangle \otimes |\psi_i\rangle \tag{3.1}$$

where i represent one of configuration of $|\Psi\rangle$ and the number of configuration is N. We might as well set $|\psi_i\rangle$ of different to be orthogonal so 3.1 is a standard schmidt decomposition form. By 1.1 we calculate the entanglement entropy of $|\Psi\rangle$ is $\log_2 N$. For a system of p particles, the possible number of configuration is 2^p and thus the entanglement entropy is $\log_2 2^p = p$ which is proportional to the size of subsystem. So we know that long-term correlation (here it is macroscopic symmetry) means entanglement volume law. Although for image set the number of cunfiguration might be way smaller than 2^p so the issue that image data set satisfies entanglement area law remains to be proved mathematically.

Another way to explore the entanglement entropy of quantum state corresponding to image data set is to explore by ansatz state of the image data set. We choose an ansatz state and utilize the data set to train the ansatz state so that it gets close to the quantum state of the image data set. After that we can calculate the entanglement entropy corresponding to the ansatz state. First exclude the form of neural network ansatz state as the entanglement entropy is also hard to calculate. There are some work on image recognisition with tensor network in recent years and the result is not good:

In 2017 (4) Stoudenmire put forward possibility of using tensor network to do supervised learning. The data set he chose is MNIST. He utilized the data set to train a couple of MPS of different bond dimension and finally get a recognition accuracy of 95% (D=10) to 99.03% (D=120). In contrast, MPS of bond dimension D=120 contains parameters of 2.5×10^7 and the performance is alike a three-layer fully-connected neural network whose recognition accuracy is 99.3%. The main reason is that MPS is of two dimensions and not suitable for it. In his second paper (5) he used Tree tensor network in replace of MPS and got a close result. The structure can give consideration on upper-lower adjacent and left-right adjacent pixels. Alike MPS, the maximal entanglement entropy TTN can represent is still bounded by its bond dimension D and is less than $\log_2 D$. But compared with MPS, TNN doesn't need all the bond dimension to reach the maximam D. The part near the root node have relatively large bond-dimension and the part near leaf node have small bond-dimension so generally speaking the effect would be better that MPS. Finally they got recognition accuracy of 98% on TNN and the largest bond dimension is 444.

While at the same time, there are other articles that used the same method with MNIST and fashion-MNIST. The results are far behind the mainstream of machine learning algorithmms. It's obvious that the data set's entanglement entropy is beyond the representation power of TNN so it's not a good idea to utilize TNN to train pictures. Thus we know the representation power is better than TNN and it's very hopeful to use neural network to do multibody calculation.

4 references

1. Quantum Tensor Networks in a Nutshell
2. Quantum neural network states: A brief review of methods and applications
3. Quantum entanglement in neural network states
4. Miles Stoudenmire E, Schwab D J. Supervised learning with quantum-inspired tensor networks
5. Stoudenmire E M. Learning relevant features of data with multi-scale tensor networks