EasyChair Preprint

№ 1239

Deep learning of matrix product states

Eric Howard

June 27, 2019

Deep learning of matrix product states

Eric Howard
Department of Physics and Astronomy
Macquarie University
Sydney, NSW, Australia
eric.howard@mq.edu.au
Centre for Quantum Dynamics
Griffith University
Brisbane, QLD, Australia

***Abstract*—Machine learning and especially deep learning architectures provide a fresh perspective on the study of many body physics phenomena. In this paper, we employ Restricted Boltzmann machines (RBM) to represent quantum many-body states and find connections that can be made useful to quantum many-body physics research, ultimately leading to a better understanding of the fundamental nature of entanglement entropy in quantum physics. In this work, we establish the conditions for translating RBMs into Matrix Product States (MPS), showing that deep learning algorithms can be exploited as a powerful tool for an efficient representation of quantum states. We present an algorithm for mapping an RBM into an MPS, with a specific proof for Ising model. We discuss the upper entropy bound and entanglement properties resulting from such a connection, together with the consequences of our results in a broader context.**

***Index Terms*—Tensor Networks, Matrix Product States, Restricted Boltzmann machines, Machine learning, Deep learning**

I. Introduction

Due to its intrinsic flexibility, deep learning [1] has found so far a wide use in several applications in physics and computer science, including computer vision, image classification, speech recognition, natural language processing or recommender systems. Machine learning and many-body physics are separate disciplines but share the common need for highly efficient representations of multivariate function classes. Big data analytics require novel methods to efficiently manage broad and diverse sets of data.

Classical information algorithms deal with pattern recognition and classification. Recently, machine learning techniques have been introduced to solve physical problems. Existing problems in machine learning, computational neuroscience, natural language processing, robotics, pattern and image recognition or signal processing generate vast amounts of high dimensional data. Tensors provide an efficient, natural and compact representation for such data via low-rank approximations. Deep learning techniques are extremely useful to compress high dimensional data into low dimensional sets and have been applied to find hidden patterns from various data input representations, with highly robust results against noise.

Restricted Boltzmann machines (RBM) [2] are one of the most applied deep learning algorithms, extremely useful in various applications requiring feature extraction or dimensional reduction methods using probability distribution modelling of diverse data input such as image, video or sound. RBMs have recently attracted great interest due to their versatility in several unsupervised and supervised learning tasks.

Recently, RBM has also attracted significant attention in the field of quantum many-body physics [3], extending the bridge between the fields of deep learning and quantum physics in general. A study of the physical rules behind neural networks could provide an answer to the great success and development of deep learning in physics.

An RBM can be seen as a bipartite undirected model describing the generative process in which a data vector can be generated from a binary hidden vector. This bipartite structure allows fast data encoding and sampling-based inference, helping with processing a great amount of data using complex models.

Previously applied to mostly data science problems, machine learning methods started to address important questions from physics, with recent advances in quantum many-body systems for reliable classification and detection of phases of matter or enhancing the quantum systems simulators.

Several schemes of machine learning, including supervised, unsupervised or reinforcement learning have been successfully applied to quantum systems of spins, bosons or fermions, in combination with gradient methods, Monte Carlo or others. Moreover, finding new methods to extract information from quantum devices has become a feasible and efficient method and important stage in the development of recent quantum simulators.

Finding efficient wavefunctions that parametrize the physically relevant degrees of freedom in an exponentially large Hilbert space is a major tasks for the quantum many body theory [4]. The main issue is the exponential complexity of quantum many-body states with the number of constituents. Highly sophisticated computational models are today capable to efficiently describe the entanglement features of many-body wave-function representations in modern condensed matter physics.

The great success behind the strong connection and exceptional overlap [5] between deep learning and physics may be buried at a more fundamental level in the laws of physics, intricately responsible for symmetry, locality, compositionality

and polynomial log-probability. An insight on quantum entanglement and its deep connection to tensor network states may deepen our understanding on symmetry, locality and causality in quantum physics.

The complexity of quantum systems is dependent on the size of the state space and grows exponentially with the number of degrees of freedom, a similar problem to the âcurse of dimensionalityâ found in machine learning. However, new machine learning methods with versatile abilities have been developed in order to recognize, classify or characterize datasets and may be used in the physical scenario under investigation.

Tensor network states have become an important tool in the study of critical systems and their dynamical features, even for long range interactions where the quantum correlations description falls beyond the area-law entanglement [6]. Tensor networks in general have been quite successful in approximating quantum states containing exponentially many parameters. A tensor network is a set of tensors with indices contracted according to a specific pattern.

Deep learning methods can be used to recognize and discover patterns in the input data. On the other hand, a study of entanglement for MPS may help with developing new deep learning algorithms [7]. Despite all achievements, there are still many challenges and concerns. For example, the underlying connection between well-known properties of classical data and entanglement properties of quantum states is still elusive.

We are searching for a constructive and close connection between RBM and MPS methods for quantum many-body systems and develop an algorithm to translate one language into another [8]. We here present a constructive connection between RBM and MPS methods. The entanglement properties of a many-body quantum system should be revealed by the dependencies required for performing a deep learning tasks.

There are several tensor networks that describe quantum many-body states as low-lying eigenstates of specific Hamiltonians, from matrix product states (MPS) and projected entangled pair states (PEPS) [9] to hierarchical structures like tree tensor networks (TTN) and multiscale entanglement renormalisation ansatz (MERA).

In particular, matrix product states (MPS) are a kind of a tensor network where the tensors are arranged in a one dimensional geometry, same as tensor train decomposition from applied mathematics. Tensor networks and MPS share many common features in their intrinsic mathematical structure. In a MPS there is one tensor per site in the quantum many-body system.

MPS can describe any state of the quantum many-body Hilbert space by increasing sufficiently the value of the bond dimension. In order to contain all states of the Hilbert space, the bond dimension size has be exponentially large. However, for one dimensional critical systems, the bond dimension diverges polynomially in size.

The Matrix product states (MPS) Ansatz describes exponentially decaying correlations reflected in the area law entanglement of the wavefunction and efficiently captures the ground states of local gapped Hamiltonians.

MPS networks are an efficient tool [10] that has been successfully applied in many-body physics and it is based on the main assumption that physically relevant many-body states do not require a description that scales exponentially with the system size, but with a polynomial number of coefficients specified. MPS is widely considered as a key tool for modelling [11]low-entangled quantum states but also used successfully to model highly-entangled states, using deep quantum circuits or high-dimensional physical systems.

MPS method is very successful in describing quantum states with relatively low entanglement entropy following the area law. The entanglement area law states that the entanglement entropy [12] of a system composed of two subsystems will scale linearly with the size of the boundary separating the two subsystems. The entanglement entropy is a measure of the information exchange between the two subsystems.

MPS have been also widely used for modelling time-evolving 1D systems, with particular applications in cold atom physics, stochastic modelling and optimization, time-periodic driving systems or quantum disordered systems. MPS methods are similar to tensor train decomposition from applied mathematics. We are interested in finding the general and optimal conditions for an MPS to efficiently translate into a specific RBM architecture as an efficient representation of quantum states.

The general goal here is to establish a proper connection between RBM and MPS methods in order to describe quantum states in statistical physics models and machine learning datasets. An RBM can be trained such that the probability distribution of the visible units can translate the probability distribution of the input data. The hidden units in the network could contain important information with significant physical meaning that can be used in pattern recognition to generate new samples from previously learned data.

The search for well defined wavefunctions is a key problem in quantum many-body physics. Here, the quantum-mechanical description of a system, including complex properties such as multi-qubit entanglement becomes mainly a data-driven problem. Machine learning architectures are able to identify phases or phase transitions in several condensed matter Hamiltonians.

Neural networks can be trained to find in raw datasets interesting order parameters or non-trivial states with no conventional order. An artificial neural network can be pictured [13] as a functional mapping of multiple variables that can be trained. In this way, many concepts and techniques from deep learning and quantum physics can be efficiently exchanged. For example, an insight into the entanglement entropy bound in MPS may quantify a number of complex features in RBM datasets.

In general, Boltzmann Machines are a particular form of log-linear Markov Random Field where the energy function is linear in its free parameters. For complex distributions, some of the variables are considered as never observed (hidden

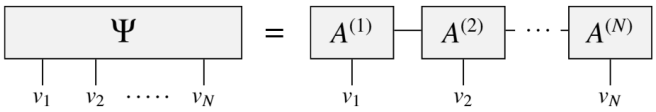

Fig. 1. The wavefunction in MPS representation

variables or units). The universal approximation theorem in machine learning states [14] that there exists an RBM to accurately describe a dataset if there is no limit on the number of hidden variables.

The modelling capacity of the Boltzmann Machine increases with the number of the hidden variables. RBMs further restrict generic Boltzmann Machines to networks without visible-visible and hidden-hidden connections. There are only connections between the visible and hidden units, not within them.

The RBMs are a specific case of energy-based models which associate a scalar energy to each configuration of the variables of interest. Learning is translated into modifying the energy function so that its shape has specific properties.

An energy-based probabilistic model defines a probability distribution via an energy function with a partition function as a normalizing factor, in analogy to physical systems [15], [16]. The energy-based model is learnt by performing stochastic gradient descent on the negative log-likelihood of the input training data.

The log-likelihood in the logistic regression will have a loss function as being the negative log-likelihood. The task is to minimise the loss function, which is done using a gradient descent algorithm. By contracting the tensor network associated with the model, the gradient of the loss can be computed.

II. RBM LEARNING OF MPS NETWORKS

If we want to describe the probability distribution of an input dataset, based on an existing model, new samples can be generated from the learned probability distribution. The data distribution can be described by a quantum state.

In a probabilistic setting, a dataset $\mathcal{T}$ can be represented by a repeated number of binary strings $\boldsymbol{v} \in \mathcal{V} = \{0,1\}^{\otimes N}$. These strings are mapped to the basis vectors of a Hilbert space of dimension 2^N. If the probability distribution is described by the wavefunction $\Psi(\boldsymbol{v})$, the collapse produced by a measurement will give a final result $\boldsymbol{v} = (v_1, v_2, \cdots, v_N)$. The probability will be proportional to $|\Psi(\boldsymbol{v})|^2$.

The quantum wavefunction using MPS is

$$\Psi(v_1, v_2, \cdots, v_N) = \left(A^{(1)v_1} A^{(2)v_2} \cdots A^{(N)v_N}\right), \quad (1)$$

where each $A^{(k)v_k}$ is a $\mathcal{D}_{k-1}$ by $\mathcal{D}_k$ matrix, and $\mathcal{D}_0 = \mathcal{D}_N$ closes the trace.

In Figure 1 the blocks represent tensors while the connected lines are tensor contractions over virtual indices. The free vertical legs are physical indices. The number of operations to be done in order to obtain the final contraction of the tensor network depends on the order [17] in which the indices are contracted. This is a significant fact and a difficult problem, because in tensor network techniques, a great number of contractions are necessary and one of the goals is to make these contractions as efficiently as possible.

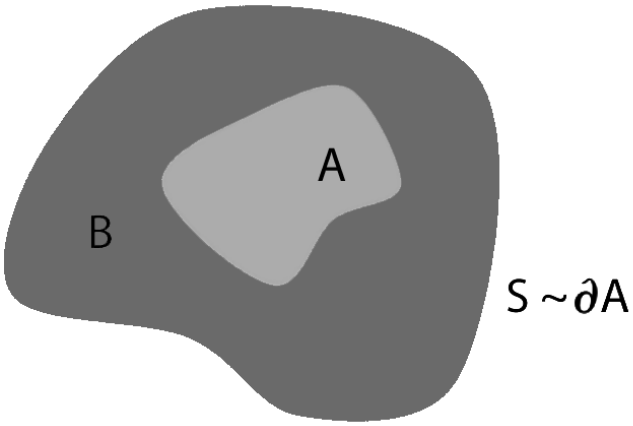

Fig. 2. The scaling of entanglement entropy in the area law

The variables are divided into two groups $\boldsymbol{v} = (\boldsymbol{v}_A, \boldsymbol{v}_B)$ and $\rho_A = \sum_{\boldsymbol{v}_B} \Psi(\boldsymbol{v}_A, \boldsymbol{v}_B)\Psi(\boldsymbol{v}'_A, \boldsymbol{v}_B)$ is the reduced density matrix one of the subsystems.

MPS can represent very efficiently a large number of quantum states [18] and accurately describe the ground states for one-dimensional gapped local Hamiltonians. The low-energy eigenstates of gapped Hamiltonians with local interactions obey the area-law for the entanglement entropy. This means that the entanglement entropy of a large system tends to scale, as the size of the boundary of the system and not as the volume. In general, the low-energy states of realistic Hamiltonians are constrained by locality so that they must obey the area law. Figure 2 shows how the entanglement entropy between two arbirary subsystems A and B scales with the size of the boundary between the two subsystems.

The Von Neumann entanglement entropy of the quantum state is $S = -(\rho_A \ln \rho_A)$. The model probability distribution can be described as

$$\mathbb{P}(\boldsymbol{v}) = \frac{|\Psi(\boldsymbol{v})|^2}{Z}, \quad (2)$$

where $Z = \sum_{\boldsymbol{v}\in\mathcal{V}} |\Psi(\boldsymbol{v})|^2$ is the normalization factor.

The RBM consists of a set of visible $v = \{v_i\}$ and hidden $h = \{h_j\}$ binary variables coupled by the coupling matrix W_{ij} and satisfying a Boltzmann distribution with the energy functional

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i W_{ij} h_j, \quad (3)$$

While the hidden units interact one with another, there is no direct interaction between the visible units here.

The number of visible and hidden binary variables are n_v and n_h. If the number of hidden units (and interactions, consequently) is increased in the RBM architecture, the network will contain functions of higher complexity between the visible

units. The biases applied to the visible and hidden units are a_i and b_j.

Due to the RBM specific structure, the visible and hidden variables are conditionally independent given one-another. If the hidden units are integrated out, the training network becomes the marginal distribution of the visible variables

$$\begin{aligned}\Psi_{\mathrm{RBM}}(v) &= \sum_h e^{-E(v,h)} \\ &= \prod_i e^{a_i v_i} \prod_j \left(1 + e^{b_j + \sum_i v_i W_{ij}}\right). \end{aligned} \quad (4)$$

The network is equivalent to a system with N qubits (spin-1/2 subsystems) described by the local basis states $|v\rangle$, where $v \in \{0, 1\}$ are the eigenstates of the z-Pauli operator $\hat{\sigma}^z |v\rangle = (-1)^v |v\rangle$. A many-body state of the system is

$$|\Psi\rangle = \sum_{\mathbf{v}} \Psi(\mathbf{v}) |\mathbf{v}\rangle, \quad (5)$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_N)^{\mathrm{T}} \in \{0, 1\}^N$ is a bit string with the configuration basis state $|\mathbf{v}\rangle$ and $\Psi(\mathbf{v})$ the associated complex amplitude.

After the MPS wavefunction is chosen, the learning process is done by adjusting the parameters of the wave function until the distribution becomes as close as possible to the data distribution. For example, the Maximum Likelihood Estimation algorithm describes a negative log-likelihood function and optimizes it by adjusting the parameters of the model.

At zero temperature, the quantum many-body problem [19], [20], [21] performs in two stages: (i) find the ground state and low-lying excitations of the Hamiltonian $\hat{H}$, and (ii) time-evolve an initial state associated with the Hamiltonian (time dependent or not). The As the quantum state possesses 2^N amplitudes $\Psi(\mathbf{v})$, the quantum many-body problem reveals a 'curse of dimensionality'. The amplitudes $\Psi(\mathbf{v})$ are represented by an order-N tensor $\Psi_{v_1 v_2 \cdots v_N}$. The main idea behind tensor networks is the decomposition of this tensor $\Psi_{v_1 v_2 \cdots v_N}$ into a network of lower order tensors.

The network is represented by a graph G where each vertex ν has an associated tensor $T^{(\nu)}$ with a small number of indices, each of dimension at most χ, with also possible physical indices v_j of dimension 2.

The edges of G describe how the internal legs of every tensor are contracted. The contraction process represents a generalisation of the matrix multiplication. A tensor network decomposition can be represented as

$$\Psi_{v_1 v_2 \cdots v_N} = \mathrm{tTr}\left[\otimes_{\nu \in G} T^{(\nu)}\right], \quad (6)$$

where tTr is the tensor trace performing the contractions of the internal indices described by G, with N open physical indices v_j. If χ scales exponentially with N, the tensor network decomposition describes any possible state $|\Psi\rangle$.

The dimension χ of the internal indices describes how much entanglement can be modelled by the network and it is illustrated by the entanglement area-law. Even for a small χ many networks describe efficiently and accurately highly compressed physical states.

For MPS models, the network essentially retains the same geometry of the underlying physical system that tries to depict, for example a chain or lattice. If the system has a coordination number $\mathcal{Z}$, the network is built from tensors with $\mathcal{Z}$ internal indices and one physical index. The RBM can be translated into an MPS by representing visible and hidden variables as physical and virtual units in the same network structure.

The process of tensor network decomposition requires to variationally minimize the tensor elements, starting with finding the expectation values $\langle\psi| \hat{h} |\psi\rangle$, using the product operator $\hat{h}$ in the Hamiltonian $\hat{H}$. This demands a contraction of the tensor network for $\langle\psi| \hat{h} |\psi\rangle$. The 1D chain intrinsic geometry makes possible the exact contraction of an MPS.

A wavefunction of n_v physical variables is represented by the MPS as

$$\Psi_{\mathrm{MPS}}(v) = \mathrm{Tr} \prod_i A^{(i)}[v_i], \quad (7)$$

where $A^{(i)}$ is a three-index tensor. For any v_i, $A^{(i)}[v_i]$ is represented by a matrix, whose dimension is the virtual bond dimension of the MPS.

In a computationally efficient description of the quantum state of a many-body system, the final representation of the wavefunction in terms of a an MPS depends on a polynomial number of parameters. If the bond dimension is increased, the MPS can better describe complex multivariable functions.

The power of an RBM depends on finding an MPS representation [22], [23], [24] with the smallest possible bond dimensions. If the network is split into n_v sites containing at least a visible variable, the contraction of the network for every site translates into merging all hidden variables. If the bond dimensions are trivial, then there is no entanglement present in the wavefunction, and the tensor network state is just a product state, as in mean field theory.

We introduce a diagonal tensor, $\Lambda_v^{(i)}$ or $\Lambda_h^{(j)}$, for every visible or hidden site and a 2×2 matrix $M^{(ij)}$ at every bond that connects each v_i to h_j

$$\Lambda_v^{(i)} = \mathrm{diag}\left(1, e^{a_i}\right), \quad (8)$$

$$\Lambda_h^{(j)} = \mathrm{diag}\left(1, e^{b_j}\right), \quad (9)$$

$$M^{(ij)} = \begin{pmatrix} 1 & 1 \\ 1 & e^{W_{ij}} \end{pmatrix}. \quad (10)$$

In this way, an MPS is generated by merging the links between sites onto virtual bonds. Overall, the MPS will posses a bond dimension that is fully dependent on the thickness of the virtual bonds and consequently on the total number of the merged links. If we have enough information about the MPS and its entanglement properties, the RBM can be efficiently generated.

In order to describe the entanglement entropy of the system, we need to divide its visible variable into two subsystems, X and Y. The entanglement entropy of a function Ψ between the two subsystems is

$$S = -\mathrm{Tr}(\rho \ln \rho), \quad (11)$$

where ρ is the reduced density matrix defined by

$$\rho = \sum_{v_Y} \Psi^* \left(v'_X, v_Y\right) \Psi \left(v_X, v_Y\right). \tag{12}$$

The entanglement entropy describing all the information from Ψ represents the short range correlations between the degrees of freedom from X and Y. v_Y contains all the visible variables in Y. The entanglement entropy depends only on the size of the boundary between the subsystems X and Y and it is bounded by the logarithm of the bond dimension, $\ln D$. If X and Y are not entangled, the entanglement entropy is zero. The maximal entanglement entropy between X and Y sites depends on the bond dimension of the MPS.

On the other hand, the RBM can be described [25], [26] by the connections between visible units via the hidden units. The geometrical structure of the hidden variables with respect to the visible ones is arbitrarily chosen. In analogy with the MPS method [27], [28], we introduce a boundary between the visible variables, splitting the system in two sites, X and $Y(= Y_1 \cup Y_2)$. The boundary Y_1 contains all visible variables connected to X, while Y_2 contains the rest of the variables.

The RBM function is

$$\Psi_{\mathrm{RBM}}\left(v\right) = \psi\left(v_X, v_{Y_1}\right) \phi\left(v_{Y_1}, v_{Y_2}\right). \tag{13}$$

If the visible variables from Y_1 are fixed, the RBM is a direct product of the visible variables from X and Y_2. The entanglement entropy between X and Y is given by the total number of visible variables at the boundary of Y_1, called $|Y_1|$. The MPS bond dimension becomes $D = 2^{|Y_1|}$.

We can also separate the region X into two sites $X = X_2 \cup X_1$ containing all variables connected to Y in X_1 and the rest of the variables in X_2. The entanglement entropy will be now bounded by the number of variables in X_1 and consequently by $S_{\max} = \min(|X_1|, |Y_1|) \ln 2$.

Any probability distribution of a N-bit system can be represented by an MPS as long as its bond dimensions are free from restrictions. As the bond dimension increases, the MPS improves its capacity to parametrize more complicated functions.

For a number of variables m, the entanglement entropy becomes

$$S_{\max} \sim m V^{(d-1)/d}, \tag{14}$$

where d is the spatial dimension, and V is the volume of the system.

More generally, the entanglement entropy of an RBM with sparse degrees of freedom follows the area law, while the entropy for dense RBMs scales with $m \sim V^{1/d}$ and consequently $S_{\max} \sim V$ grows linearly with the volume. As the number of parameters for a dense RBM scales polynomially with the full size of the system, these RBMs can describe highly entangled quantum states where the area law is not satisfied.

The connecting bond indices (or ancillary indices) describe the structure of the many-body entanglement in the quantum state. If the bond dimension is changed, the multiplicative factor of the area law also changes. If we want to modify the scaling with the boundary size, the geometric pattern of the tensor network has to change. The entanglement in the tensor network is a consequence of both of the bond dimension and the geometric pattern of the interconnection between bond indices. The number of the values that any of the bond indices can take is a measure of the amount of the quantum correlations encoded in the wavefunction.

The correlation functions of the MPS decay exponentially with the correlation distance. The correlation length of MPS states is finite, therefore MPS can't fully describe the properties of critical or scale-invariant systems, where the correlation length diverges.

The condition for an MPS to posses an RBM representation is that each tensor has the form

$$A_{\alpha\beta}\left[v\right] = L_{\alpha v} R_{v\beta}, \tag{15}$$

where L and R are two 2 x 2 matrices. The product of R and L is described by the coupling to the hidden RBM variables. The product RL will take the symmetric form

$$\begin{aligned} RL \;=\; & \begin{pmatrix} 1 & \\ & e^{a/2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & e^{W} \end{pmatrix} \begin{pmatrix} 1 & \\ & e^{b} \end{pmatrix} \\ & \begin{pmatrix} 1 & 1 \\ 1 & e^{W} \end{pmatrix} \begin{pmatrix} 1 & \\ & e^{a/2} \end{pmatrix}. \end{aligned} \tag{16}$$

As an MPS representation has gauge degrees of freedom, this this makes possible to restrict the tensors with canonical conditions, with direct benefits on computing an exact partition function. As an example, the partition function of the statistical Ising model with $s_i = \pm 1$ Ising spins is

$$Z = \sum_{\{s_i\}} \exp\left(K \sum_{\langle i,j \rangle} s_i s_j + H \sum_i s_i\right) \tag{17}$$

where K is the coupling constant, H the external field and the binary variables are defined as $v_i = (s_i + 1)/2$.

The wavefunction can be complex valued, but here we restrict it to be real valued. The partition function can be represented as a summation of the RBM function. In more physical terms, the RBM amplitudes can be translated into the effective action for physical spins which is obtained by tracing out a bath of hidden spin variables.

For the one dimensional Ising case, the partition function can be represented as an MPS. The matrix product on each bond is

$$RL = \begin{pmatrix} e^{K+H} & e^{-K} \\ e^{-K} & e^{K-H} \end{pmatrix}. \tag{18}$$

III. Conclusion

We studied the close connection between the RBM and MPS and presented a simple algorithm to transform an RBM into an MPS. This connection may may open the door into applying methods developed in quantum physics directly into machine learning. A condition for finding an RBM representation for a given MPS was also found, together with an upper bound on the entanglement entropy of a quantum system modelled as an parametrized RBM.

We also discussed the equivalence of a number of features and their entanglement properties for RBM and MPS, quantifying the capacity of RBMs to help with a better understanding of entanglement theory in the classical setting for quantum many-body physics.

There is a visible crossover from classically simulable tensor networks to algorithms requiring a quantum computer to evaluate. RBM networks already provide good results for a large set of learning tasks using classical resources, with no requirement for quantum hardware.

As the matrix product intrinsic structure has the capability for many operations to be efficiently executed at a classical level, a bridge between RBMs and MPS models can further improve and accelerate the algorithmic development of hybrid quantum/classical subroutines and the use of classical algorithms on quantum computers.

REFERENCES

[1] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", Nature **521**, 436-444 (2015).

[2] N. Le Roux, Y. Bengio, "Representational Power of Restricted Boltzmann Machines and Deep Belief Networks", Neural Comput. **20**, 1631-1649 (2008).

[3] G. Carleo, M. Troyer, "Solving the quantum many-body problem with artificial neural networks", Science **355**, 602 (2017).

[4] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, "Restricted Boltzmann machine learning for solving strongly correlated quantum systems", Phys. Rev. B **96**, 205152 (2017).

[5] H. Saito and M. Kato, "Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice", J. Phys. Soc. Jpn. **87**, 014001 (2018).

[6] L. Amico, R. Fazio, A. Osterloh, V. Vedral, "Entanglement in Many-Body Systems", Rev. Mod. Phys. **80**, 517 (2008).

[7] X. Gao, L.-M. Duan, "Efficient representation of quantum many-body states with deep neural networks", Nature Communications **8**, 662 (2017).

[8] G. Vidal, "Class of Quantum Many-Body States that Can Be Efficiently Simulated", Phys. Rev. Lett **101**, 110501 (2008).

[9] F. Verstraete, V. Murg, J.I. Cirac, "Matrix Product States, Projected Entangled Pair States, and Variational Renormalization Group Methods for Quantum Spin Systems", Advances in Physics, **57** (2), 143-224 (2008).

[10] L. Tagliacozzo, G. Evenbly, G. Vidal, "Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law", Phys. Rev. B **80**, 235127 (2009).

[11] M. B. Hastings, "An Area Law for One-Dimensional Quantum Systems", J. Stat. Mech. (2007) P08024.

[12] G. Vidal, J. I. Latorre, E. Rico, A. Kitaev, "Entanglement in Quantum Critical Phenomena", Phys. Rev. Lett **90**, 227902 (2003).

[13] G. Carleo, Y. Nomura, M. Imada, "Constructing exact representations of quantum many-body systems with deep neural networks", arXiv:1802.09558 (2018).

[14] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, "Equivalence of restricted Boltzmann machines and tensor network states", Phys. Rev. B **97**, 085104 (2018).

[15] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent, Proceedings of the 19th International Conference on Computational Statistics", 177-187 (2010).

[16] D. P. Kingma, J. L. Ba, "Adam: A method for stochastic optimization", arXiv:1412.6980 (2014).

[17] J. Eisert, M. Cramer, M. B. Plenio, "Colloquium: Area Laws for the Entanglement Entropy", Rev. Mod. Phys. **82**, 277 (2010).

[18] I. P. McCulloch, "Infinite size density matrix renormalization group", revisited, arXiv:0804.2509 (2008).

[19] G. Vidal, "Efficient Simulation of One-Dimensional Quantum Many-Body Systems", Phys. Rev. Lett **93** (4), 040502 (2004).

[20] R. Jastrow, "Many-Body Problem with Strong Forces", Phys. Rev. **98** (5), 1479-1484 (1955).

[21] M. C. Gutzwiller, "Correlation of Electrons in a Narrow s Band", Phys. Rev. **137** (6A), A1726-A1735 (1965).

[22] Y. Huang, J. E. Moore, "Neural network representation of tensor network and chiral states", arXiv:1701.06246 (2017).

[23] S. R. Clark, "Unifying Neural-network Quantum States and Correlator Product States via Tensor Networks", J. Phys. A: Math. Theor. **51**, 135301 (2018).

[24] L. Bottou, F. E. Curtis, J. Nocedal, "Optimization Methods for Large-Scale Machine Learning", Siam Reviews, **60** (2), 223-311 (2018).

[25] J. Carrasquilla, R. G. Melko, "Machine Learning Phases of Matter", Nature Physics **13**, 431-434 (2017).

[26] L. Wang, "Discovering Phase Transitions with Unsupervised Learning", Phys. Rev. B **94**, 195105 (2016).

[27] D.-L. Deng, X. Li, S. Das Sarma, "Quantum Entanglement in Neural Network States", Phys. Rev. X **7**, 021021 (2017).

[28] D.-L. Deng, X. Li, S. Das Sarma, "Machine learning topological states", Phys. Rev. B **96**, 195145 (2017).