Unsupervised Recognition of Informative Features via Tensor Network
Machine Learning and Quantum Entanglement Variations

Sheng-Chen Bai(白生辰)†, Yi-Cheng Tang(唐以成)†, and Shi-Ju Ran(冉仕举)*

Department of Physics, Capital Normal University, Beijing 100048, China

(Received 8 August 2022; accepted manuscript online 15 September 2022)

Given an image of a white shoe drawn on a blackboard, how are the white pixels deemed (say by human minds) to be informative for recognizing the shoe without any labeling information on the pixels? Here we investigate such a “white shoe” recognition problem from the perspective of tensor network (TN) machine learning and quantum entanglement. Utilizing a generative TN that captures the probability distribution of the features as quantum amplitudes, we propose an unsupervised recognition scheme of informative features with variations of entanglement entropy (EE) caused by designed measurements. In this way, a given sample, where the values of its features are statistically meaningless, is mapped to the variations of EE that statistically characterize the gain of information. We show that the EE variations identify the features that are critical to recognize this specific sample, and the EE itself reveals the information distribution of the probabilities represented by the TN model. The signs of the variations further reveal the entanglement structures among the features. We test the validity of our scheme on a toy dataset of strip images, the MNIST dataset of hand-drawn digits, the fashion-MNIST dataset of the pictures of fashion articles, and the images of nerve cord. Our scheme opens the avenue to the quantum-inspired and interpreted unsupervised learning, which can be applied to, e.g., image segmentation and object detection.

DOI: 10.1088/0256-307X/39/10/100701

Machine learning (ML) such as deep learning has gained tremendous successes in an extremely wide range of fields such as computer vision and natural language processing. Such methods have strong demands on the labeled samples to extract useful information in a data-driven manner. However, labeled data are rare in many scenarios such as scientific images. Exploring efficient and reliable schemes for the unsupervised[1] and few-shot[2] learning is at the cutting edges of ML and artificial intelligence.

This paper focuses on the connection between the statistics of the model and the interpretability[4] of the model. For traditional ML, there are many challenges[4] in interpretability. Among them, a promising pathway to the unsupervised learning is to develop interpretable “white-box” ML schemes[5] by cooperating with the probabilistic theories and models, where we have for instance the information bottleneck theory[6] and Bayesian inference.[7]

In recent years, tensor network (TN), which originated from the fields of quantum physics,[8–12] sheds light on the novel quantum ML schemes[13] interpreted by quantum probabilistic theories and quantum many-body physics. TN ML has been successfully applied to the supervised, unsupervised, and reinforcement learning for various tasks including classification,[14–16] generation,[19–21] feature selection,[22] compressed sampling,[23] and anomaly detection,[24] etc. Experiments of running TN ML on quantum hardware are also in hot debate.[25,26]

In this work, we propose to unsupervisedly recognize the informative features via the generative TN[19] and its entanglement entropy (EE).[22] Given an image, an ML model simply sees a bunch of numbers (pixels) that are statistically meaningless. However, a human mind can easily recognize the critical pixels for identifying the content of the image without explicitly learning any labeling information on the pixels. Taking the three images on the left-hand side of Fig.1 as examples, a human mind could easily recognize the white pixels picturing the objects (strip, “2”, and shoe). To seek for a mathematically understandable and modeling of such recognition, we suggest to map the pixels of a given image to the statistically meaningful quantities by the single-qubit measurements on a generative TN according to the pixels of this image. Specifically, we propose to use the average variations of EE (denoted as $\langle \delta S \rangle_m$ for the $m$-th pixel) for unsupervised feature selection. The pixels with large $\langle \delta S \rangle_m$ (dubbed as the critical minority) outline the critical shape for recognizing this image, coinciding with the human minds. We test the proposed method on a toy dataset of strips, the MNIST[27] and fashion-MNIST[28] datasets. Our scheme differs from the existing unsupervised feature selection methods,[29] such as the filter methods (see, e.g., Refs. [30,31]) and the clustering and dimensionality reduction methods (e.g., Refs. [32–34]). These methods require multiple samples and their processings. We finally apply our method for image segmentation provided with just one image of nerve cord[35] without any labeling information, and raise the open question on generalizing to the feature...

†They contributed equally to this work.
*Corresponding author. Email: sjran@cnu.edu.cn
© 2022 Chinese Physical Society and IOP Publishing Ltd
selection by multi-qubit measurements.

![Diagram](image)

**Fig. 1.** Provided with the images of a vertical strip, a digit “2” and a shoe drawn on the black background (left-hand-side), a human mind can easily recognize the shapes of the images as the white pixels, which are dubbed as the critical minorities (right-hand-side), without explicitly learning any labeling information on the pixels. We propose an unsupervised TN scheme to identify the critical minority by the variations of the entanglement entropies.

**Mapping to the Variations of Entanglement Entropies via Generative Tensor Network and Measurements.** The first step of modeling the probability distribution by a generative TN is to map the samples to the quantum Hilbert space (known as the feature map\(^{[14]}\)) as

\[
\mathbf{v}^{[n]} = \prod_{\alpha = 1}^{M} \left[ \cos \left( \frac{x_{\alpha}^{[n]} \pi}{4} \right), \sin \left( \frac{x_{\alpha}^{[n]} \pi}{4} \right) \right]^T,
\]

with \(x_{\alpha}^{[n]} = [x_{1}^{[n]}, \ldots, x_{M}^{[n]}]\) the \(n\)-th sample consisting of \(M\) features.\(^{[36]}\) One can see that \(\mathbf{v}^{[n]}\) is an \(M\)-th order tensor or a \(2^M\)-dimensional vector. Obviously, \(\mathbf{v}^{[n]}\) is normalized satisfying \(\|\mathbf{v}^{[n]}\| = 1\) (L2 norm), and can be considered as the coefficient vector of an \(M\)-qubit quantum product state.

Considering the generative model as a normalized \(M\)-th order tensor \(\Psi\), the probability of generating a specific sample \(\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_M)\) follows Born’s probabilistic interpretation of quantum mechanics with

\[
P(\tilde{x}) = \left| \sum_{s_1 \ldots s_M} \Psi_{s_1 \ldots s_M} \tilde{\psi}_{s_1 \ldots s_M} \right|^2,
\]

with \(\tilde{\psi}\) defined by Eq. (1) with \(\tilde{x}\). \(\Psi\) can be regarded in general as an \(M\)-qubit entangled state. A generative TN is trained so that the probability of generating each sample \(\tilde{x}\) in the training set approaches \(1/N\) with \(N\) the total number of training samples. To this end, Han et al.\(^{[19]}\) proposed to write \(\Psi\) into a widely used TN, namely matrix product state (MPS),\(^{[8,10,37]}\) also known as the tensor-train form\(^{[38]}\), which is formed by \(M\) local tensors \(\{A^{[m]}\}\) as

\[
\Psi_{s_1 \ldots s_M} = \sum_{a_1 \ldots a_{M-1}} A^{[1]}_{s_1 a_1} A^{[2]}_{a_1 a_2 a_3} \ldots A^{[M-1]}_{a_{M-1} a_M} A^{[M]}_{s_M a_M},
\]

Here, \(\{s_m\}\) are called the physical indexes, and \(\{a_m\}\) the virtual indexes whose dimension (denoted as \(\chi\)) is a hyper-parameter that controls the parameter complexity of the MPS. A sweep algorithm\(^{[14]}\) inspired by the density matrix renormalization group\(^{[19,40]}\) is used to optimize the local tensors \(\{A^{[m]}\}\) to minimize the following loss:

\[
L = -\frac{1}{N} \sum_{n=1}^{N} \ln P(x^{[n]}).
\]

To explain the main idea of this work, we design a toy dataset of vertical strips [Fig. 2(a)]. The whole region consists of three parts [Fig. 2(b)]. In the outer rim, the pixels are taken to be white (with \(x_m = 0\)) for all samples, thus contain no information at all, which we name as the background. A vertical strip (with \(x_m = 0.1\)) appears at different positions in the black square region (with \(x_m = 1\)) in the middle, which we dub as the informative area. Particularly, the pixels of the strip in each image are referred as the critical minority that we assume to carry the most critical information of the image. This is a reasonable assumption as a human, after briefly reading the images in this dataset, could easily recognize the “moving” strips as the critical minority.

![Diagram](image)

**Fig. 2.** (a) Several samples in a toy dataset of vertical strips, and (b) the illustrations of the background, informative area, and critical minority.

The main points of this work are as follows:

- The EE of the MPS indicates the informative area and the background for the dataset (which is in general sample-independent, similar to the existing unsupervised feature selection methods\(^{[29]}\)).
- The average variations of the EE by measuring the MPS indicate the critical minority of a specific image (which is sample-dependent).

Figure 3 take two samples (see the first column) as examples to demonstrate the entanglement information obtained from the generative MPS. The second column shows the EE of MPS. The EE corresponding to the \(m\)-th pixel (or physical index) is defined as

\[
S_{m} = -\text{Tr} \left( \rho^{[m]} \ln \rho^{[m]} \right),
\]

where \(\rho^{[m]}_{s_m} = \sum_{s_m} \Psi_{s_1 \ldots s_{M-2} s_m s_{M-1}} \Psi_{s_1 \ldots s_{M-2} s_m s_{M-1}}\) is the reduced matrix of the \(m\)-th physical index (\(\sum_{s_m}\) means to sum over all but the \(m\)-th physical indexes). The EE of a single qubit characterizes the amount of uncertainty that can be reduced by measuring this qubit. Thus, a larger EE indicates more information obtained by measuring this.
qubit, and *vice versa*. In the background, the EE is zero as expected since mathematically the corresponding qubits form the unentangled product states for the strip dataset. The EE in the informative area is approximately uniform with $S_m \simeq 0.2$, since the probabilities of having a strip at different positions are uniform. The distribution of the EE clearly identifies the pixels that contain non-trivial information.

**Fig. 3.** Taking two images (first column) from the dataset of strips to show the EE of the generative MPS [second column; see Eq. (5)] and the average variations of EE [third column; see Eq. (7)] by measurements according to the images.

To identify the informative features that are critical for a specific sample, we investigate the variations of the EE by measuring on one qubit according to the value of the corresponding feature. Given a specific sample $\tilde{x}$, we measure on the $m'$-th qubit (i.e. the eigenvector at the $m'$-th position was contracted with MPS) and have

$$\phi_{s_1 \ldots s_{m'-1} s_{m'} 1 \ldots s_M}^{[m']} = \sum_{s_{m'}} \psi_{s_1 \ldots s_M}^{[m']}$$

with the vector $\psi^{[m']}$ obtained using the map on the $m'$-th feature $\tilde{x}_{m'}$ of the given image. By normalizing $\phi^{[m']}/|\phi^{[m']}| \rightarrow \Phi^{[m']}$, it represents an $(M - 1)$-qubit state that captures the posterior probability distribution of the $(M - 1)$ unmeasured features in the condition of knowing $\tilde{x}_{m'}$. The average variation of the EE after the measurement is defined as

$$\langle \delta S \rangle_{m'} = \sum_{m \neq m'} \frac{(S_m - S_{m'})}{M - 1},$$

with $S_m$ and $S_{m'}$ the EE of the $m$-th qubit before and after the measurement, respectively. In short, Eq. (7) maps a given sample of $M$ features to $\langle \delta S \rangle_{m'} (m' = 1, \ldots, M)$.

The third column of Fig. 3 shows the $\langle \delta S \rangle_{m'} (m' = 0, \ldots, M)$ by measuring each qubit respectively according to the image in the given first column. For the critical minority (the pixels of the white strip), we clearly obtain much larger EE variations with $\langle \delta S \rangle_{m'} \sim O(10^{-4})$. In the informative area but outside the strip, we have $\langle \delta S \rangle_{m'} \sim O(10^{-6})$, which are nonzero but much smaller than those of the critical minority. For the background, we have $\langle \delta S \rangle_{m} = 0$ since the EEs before and after the measurement are zero. Our results show that $\langle \delta S \rangle_{m'}$ can mark the critical minority fairly well, though we do not have any prior information on labeling the pixels. The larger EE variations in the strip are essentially due to the fact that the black pixels in the strip are a monitor compared with the rest ones within the informative area. Consequently, knowing a pixel to be white (in the informative area) will largely reduce the EE of the qubits in the same column by knowing the position of the strip. In comparison, knowing a pixel to be black only excludes this column as the position of the strip, where the decrease of the EE should be relatively small. The generative MPS captures such properties in a simple manner: the qubits are more entangled strip-wisely.

To provide an intuitive understanding, let us consider the following three-qubit state as a simplest example:

$$\Psi = \frac{1}{\sqrt{2}} [1 \otimes (1 \otimes 0)] + [0 \otimes 1 \otimes 0]. \quad (8)$$

This state can be considered to describe the probability distribution of two samples $x^{[1]} = (0, 0, 1)$ and $x^{[2]} = (0, 1, 0)$, with $P(x^{[1]}) = P(x^{[2]}) = 0.5$. The last two qubits form a maximally entangled “singlet” state. From Eq. (5), we have $S = 0$ for the first qubit, and $S = \ln 2$ for the last two. The first qubit can be recognized as the background.

Considering a specific sample $\tilde{x}$ with knowing $\tilde{x}_1 = 0$, we accordingly measure on the third physics index of $\Psi$ in Eq. (8) by following Eq. (6), and have $\Phi^{[3]} = [1, 0]^T \otimes [0, 1]^T$, with $S = 0$ for the rest two qubits. The average variations of EE $\langle \delta S \rangle_1 = \langle \delta S \rangle_2 = (0 - \ln 2)/2 = -(\ln 2)/2$ are negative. One can see that the last two qubits are highly entangled before the measurement, similar to the qubits in a same vertical strip. The measurement (in the basis of $[0, 1]^T$ and $[1, 0]^T$ in this case) on one qubit will generally eliminate the uncertainty of the other. This simple example implies that in the more complicated cases, a negative $\langle \delta S \rangle_{m'}$ may suggest a relatively large entanglement between the measured qubit and (some of) the rest. The measurement on such a qubit (in other words, under the condition of knowing the value of the corresponding feature) will result in a probability distribution with smaller uncertainty. Obviously, the same discussions can be made if we reverse the black and white colors in the images.

**Testing on Sophisticated Datasets.** We test our scheme on more sophisticated datasets, which are the MNIST dataset with the images of hand-drawn digits, fashion-MNIST dataset with the images of articles, and the images of nerve cord. For each class in a dataset, we train a generative MPS for evaluating the entanglement properties. We take four training samples as examples shown in the first column of Fig. 4. The second column demonstrates the $S$ of the MPSs. The relatively large EE (red regions) marks the informative areas that approximately form the shapes of the corresponding digits or articles. Note again the informative areas are from the properties of the generative MPSs, thus do not depend on any specific samples. For instance, the last two sub-figures in the second column are the same, showing the EE of the generative MPS for shoes.

The third column shows the average EE variations $\langle \delta S \rangle_{m'}$, which identify the critical monitores of the specific images shown in the first column. The distinct shapes
of the original images are successfully outlined by $\langle \delta S \rangle_{m'}$. For instance, the special writing habit in the “2”, the rectangle printed on the T-shirt, and the different styles of the shoes are reflected by $\langle \delta S \rangle_{m'}$, which do not appear in the illustrations of the $S$ of the MPS shown in the middle column. Particularly, the critical minorities of the two shoes are obtained from the same generative MPS, and the distinct shapes of these two images are still well identified by $\langle \delta S \rangle_{m'}$.

$$\delta S = \ln 2 > 0.6365.$$  

The EE increases after the measurement with $\langle \delta S \rangle > 0$. In this case, the probability of $P(x_k = \tilde{x}_k)$ is in general small (notice that $x_k$ denotes the feature corresponding to the measured qubit and $\tilde{x}_k$ denotes the value of this feature in the specific sample). The measurement will (relatively) largely enhance the probabilities of the samples that also have $x_k = \tilde{x}_k$, which may be small before the measurement. Consequently, the uncertainty of the unmeasured qubits may increase, leading to $\langle \delta S \rangle > 0$.

$$\Psi = \frac{1}{\sqrt{3}} \left( \begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right) \otimes \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \otimes \left( \begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right) + \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \otimes \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \otimes \left( \begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right).$$

This describes the probability distribution of three samples $x^{[1]} = (0, 1, 0)$, $x^{[2]} = (1, 0, 0)$, and $x^{[3]} = (1, 1, 1)$ with identical probabilities. All the three qubits are entangled, where the EE of the first two qubits is $S \simeq 0.6365$. Consider again a sample $\tilde{x}$ with $\tilde{x}_3 = 0$. By measuring on the third qubit accordingly, the first two qubits will be projected into a maximally entangled state with $S = \ln 2 > 0.6365$.
tured data, where the importance of the original features exhibits obvious difference. A promising way of dealing with the unstructured data is to combine with the feature extraction methods. Further improvement by generalizing from single-qubit to multi-qubit measurements in the definition of the entanglement variations (see some preliminary results in the Supplemental Material) are also to be investigated in the future for different kinds of ML tasks.

Acknowledgment. We are indebted to Yuhan Liu, Zheng-Zhi Sun, Ke Li, Peng-Fei Zhou, Rui Hong, Jia-Hao Wang, and Wei-Ming Li for helpful discussions. This work was supported by the Chinese Natural Science Foundation of China (Grant Nos. 12004266 and 11834014) and the Foundation of Beijing Education Committees (Grant No. KM202010028013). SJR acknowledges the support from the Academy for Multidisciplinary Studies, Capital Normal University.

References

[1] Barlow H B 1989 Neural Comput. 1 295
[2] Wang Y, Yao Q, Kwok J T, and Ni L M 2020 ACM Comput. Surv. 53 63
[3] Molnar C 2022 Interpretable Machine Learning 2nd edn (Osano, Inc., A Public Benefit Corporation)
[4] Rudin C, Chen C, Chen Z, Huang H, Semenova L, and Zhong C 2022 Stat. Surveys 16 1
[5] Gilpin L H, Bau D, Yuan B Z, Bajwa A, Specter M, and Gordon, Greenspan, and Goldberger 2003 Proceedings Ninth IEEE International Conference on Data Science and Advanced Analytics (DSAA) pp 80–89
[6] Gordon, Greenspan, and Goldberger 2003 Proceedings Ninth IEEE International Conference on Computer Vision pp 370–377
[7] Li F F, Fergus R, and Perona P 2003 Proceedings Ninth IEEE International Conference on Computer Vision pp 1134–1144
[8] Verstraete F, Murg V, and Cirac J I 2008 Adv. Phys. 57 143
[9] Cirac J I and Verstraete F 2009 J. Phys. A: Math. Theor. 42 504004
[10] Ran S J, Tirrito E, Peng C, Chen X, Tagliacozzo L, Su G, and Lewenstein M 2020 Tensor Network Contractions: Methods and Applications to Quantum Many-Body Systems (Berlin: Springer)
[11] Orús R 2019 Nat. Rev. Phys. 1 538
[12] Cirac J I, P´erez-Garc´ıa D, Schuch N, and Verstraete F 2021 Rev. Mod. Phys. 93 045003
[13] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, and Lloyd S 2017 Nature 549 195
[14] Stoudenmire E and Schwab D J 2016 Advances in Neural Information Processing Systems 29 (Curran Associates, Inc.) pp 4799–4807
[15] Liu D, Ran S J, Wittek P, Peng C, Garc´ıa R B, Su G, and Lewenstein M 2019 New J. Phys. 21 073059
[16] Sun Z Z, Peng C, Liu D, Ran S J, and Su G 2020 Phys. Rev. B 101 075135
[17] Cheng S, Wang L, and Zhang P 2021 Phys. Rev. B 103 125117
[18] Stoudenmire E M 2018 Quantum Sci. Technol. 3 034003
[19] Han Z Y, Wang J, Fan H, Wang L, and Zhang P 2020 Phys. Rev. X 8 031012
[20] Cheng S, Wang L, Xiang T, and Zhang P 2019 Phys. Rev. B 99 155131
[21] Viejra T, Vanderstraeten L, and Verstraete F 2022 arXiv:2202.08177 [quant-ph]
[22] Liu Y, Li W J, Zhang X, Lewenstein M, Su G, and Ran S J 2021 Front. Appl. Math. Stat. 7 716044
[23] Ran S J, Sun Z Z, Fei S M, Su G, and Lewenstein M 2020 Phys. Rev. Res. 2 033293
[24] Wang J, Roberts C, Vidal G, and Leichenauer S 2020 arXiv:2006.02516
[25] Wang K, Xiao L, Yi W, Ran S J, and Xue P 2021 Photon. Res. 9 2332
[26] Wall M L, Abernathy M R, and Quiriz G 2021 Phys. Rev. Res. 3 023010
[27] Deng L 2012 IEEE Signal Process. Mag. 29 141
[28] Xiao H, Rasul K, and Vollgraf R 2017 arXiv:1708.07747 [cs.LG]
[29] Solatorio-Fernández S, Carrasco-Ochoa J A, and Martínez-Trinidad J F 2020 Artificial Intell. Rev. 53 907
[30] Varshavsky R, Gottlieb A, Linial M, and Horn D 2006 Bioinformatics 22 e507
[31] Tabakhi S, Najafi A, Ranjbar R, and Moradi P 2015 Neurocomputing 168 1024
[32] Dy J G and Brodley C E 2004 J. Mach. Learn. Res. 5 845
[33] Kim S B and Rattakorn P 2011 PLoS ONE 6 e19451
[34] Yao J, Mao Q, Goodison S, Mai V, and Sun Y 2015 Expert Syst. Appl. 42 504004
[35] Cardona A, Saalfeld S, Preibisch S, Schmid B, Cheng A, Patan´e A, and Chang T 2018 Bioinformatics 34 3704
[36] Varshavsky R, Gottlieb A, Linial M, and Horn D 2006 Bioinformatics 22 e507
[37] Tabakhi S, Najafi A, Ranjbar R, and Moradi P 2015 Neurocomputing 168 1024
[38] Dy J G and Brodley C E 2004 J. Mach. Learn. Res. 5 845
[39] Kim S B and Rattakorn P 2011 Expert Syst. Appl. 38 5704
[40] Yao J, Mao Q, Goodison S, Mai V, and Sun Y 2015 Pattern Recognt. Lett. 53 100
[41] Cardona A, Saalfeld S, Preibisch S, Schmid B, Cheng A, Pulokas J, Tomancak P, and Hartenstein V 2010 PLoS Biology 8 e1000502
[42] Here we avoid to use Dirac’s symbols for readers who are not familiar with quantum physics. We use a bold letter to represent a tensor (or matrix, vector, such as $\Phi$), and the same normal letter with lower indexes to represent the tensor elements (such as $\Phi_{1 \ldots N}$).
[43] Pérez-García D, Verstraete F, Wolf M M, and Cirac J I 2007 Quantum Inf. Comput. 7 401
[44] Oseledets I V 2011 SIAM J. Sci. Comput. 33 2295
[45] White S R 1992 Phys. Rev. Lett. 69 2863
[46] White S R 1993 Phys. Rev. B 48 10345

100701-5