Matrix Product State Curriculum Learning

Hyun Ho Cha, Jung Woo Lee*

Seoul National University

aiden132435@cml.snu.ac.kr, *junglee@snu.ac.kr

Abstract

A curriculum learning framework for Matrix Product State (MPS) models that accelerates training on high-dimensional data is proposed. Starting from low-dimensional inputs, we train a shorter MPS of length N by replacing consecutive site-wise contraction with matrix power operations. A Z-order curve preserves spatial locality when linearizing images into the tensor chain. As the curriculum advances, we gradually reduce the pooling window (or exponent) until the full-resolution MPS is recovered. Experiments on MNIST show faster convergence and comparable final accuracy.

Ⅰ. Introduction

In modern machine learning, high-dimensional data such as images or time series often exhibit strong local correlations that can be exploited to reduce model complexity [1]. One promising approach is to represent such data sequentially and process it with tensor network models originally developed in quantum many-body physics. In particular, the Matrix Product State (MPS) architecture [2, 3] provides a highly parameter-efficient ansatz by decomposing a full tensor of order N into a chain of low-order tensors interconnected by moderate-size "bond" dimensions [4].

Despite their representational power, training an MPS on full-resolution inputs can be expensive: the number of tensor contractions grows linearly in N, and the optimization landscape becomes more complex as the chain length increases. To mitigate these issues, we introduce a curriculum learning strategy that progressively increases input resolution during training, allowing the model to first capture coarse-grained structure before refining fine details.

A tensor of order d is a multidimensional array

$$T\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_d},$$

whose entries we denote by $T_{i_1 i_2 \cdots i_d}$, where each index i_k runs from 1 to n_k. A contraction of two tensors

$$T\in\mathbb{R}^{\cdots\times a\times\cdots} \text{ and } S\in\mathbb{R}^{\cdots\times a\times\cdots}$$

over a common index of size a is defined by summing over that index. For example, it $T\in\mathbb{R}^{m\times a}$ and $S\in\mathbb{R}^{a\times n}$, their contraction over the shared index of size a is the matrix product $(T\cdot S)\in\mathbb{R}^{m\times n}$, where

$$(T\cdot S)_{ij}=\sum_{k=1}^{a}T_{ik}S_{kj}.$$

Scalar regression can be performed by an MPS, which consists of N third-order tensors $A^{(1)},A^{(2)},\ldots,A^{(N)}$, where each

$$A^{(k)}\in\mathbb{R}^{m_{k-1}\times 2\times m_k},$$

with bond dimensions $m_0=m_N=1$ and $m_k=m$ for $1\le k\le N-1$. We index the modes of $A^{(k)}$ as $A^{(k)}_{i_{k-1}j_k i_k}$. Given a one-dimensional input $x\in[0,1]$, we encode it into a "physical" vector

$$\phi(x)=\begin{pmatrix}\cos\left(\frac{\pi}{2}x\right)\\ \sin\left(\frac{\pi}{2}x\right)\end{pmatrix}\in\mathbb{R}^2,$$

and write its components as $\phi_{j_k}(x_k)$ if x_k is the kth pixel in a sequence. The scalar output of the MPS is then the full contraction

$$f(x_1,\ldots,x_N)=\sum_{i_0,\ldots,i_N}\sum_{j_1,\ldots,j_N}\left[\prod_{k=1}^{N}A^{(k)}_{i_{k-1}j_k i_k}\phi_{j_k}(x_k)\right],$$

which yields a single real number since $i_0=i_N=1$.

This can be extended to L-class classification. To produce an L-dimensional output vector $y\in\mathbb{R}^L$ for classification, we promote exactly one of the site-tensors—say the Kth tensor—to fourth order:

$$A^{(K)}\to\tilde{A}^{(K)}\in\mathbb{R}^{m_{K-1}\times 2\times m_K\times L},$$

with entries $\tilde{A}^{(K)}_{i_{K-1}j_K i_K l}$, $l=1,\ldots,L$. All other tensors remain third order. The lth component of the MPS output is

$$y_l(x_1,\ldots,x_N)=\sum_{\substack{i_0,\ldots,i_N\\ j_1,\ldots,j_N}}\left(\prod_{\substack{k=1\\ k\neq K}}^{N}A^{(k)}_{i_{k-1}j_k i_k}\phi_{j_k}(x_k)\right)\times\tilde{A}^{(K)}_{i_{K-1}j_K i_K l}\phi_{j_K}(x_K).$$

Stacking all L outputs gives the vector

$$y(x_1,\ldots,x_N)=[y_1,y_2,\ldots,y_L]^T\in\mathbb{R}^L,$$

which can then be passed through a softmax if desired for classification probabilities.

Ⅱ. Method

In the MPS-based classifier, the full-resolution input has length $N=28\times 28=784$, so a direct training of an MPS of length N can be computationally demanding. To accelerate learning, we adopt a *curriculum learning* strategy [5] in which we gradually increase the input resolution. Concretely, we first apply average pooling with window size $H\times W$ to each 28×28 image, so that each block of $H\times W$ pixels is replaced by its mean intensity. The pooled image then has size $(28/H)\times(28/W)$, and we train a corresponding MPS of length $N_{\text{low}}=784/HW$. Although this low-resolution MPS has only N_{low} sites, we would like its contractions to mimic those of the full-resolution model. Let $A^{(k)}$ be the kth MPS tensor and $\phi(x)\in\mathbb{R}^2$ the local feature map for a pooled pixel x. Normally the contraction

at site k is $A^{(k)} \cdot \phi(x) \in \mathbb{R}^{m_{k-1} \times m_k}$. To simulate the effect of having HW consecutive pixels at full resolution, we instead take the matrix power $\left(A^{(k)} \cdot \phi(x)\right)^{HW}$, i.e., raise the site-matrix to the HW th power. This "exponentiated" contraction embeds the contributions of HW identical pixels into a single step, without explicitly unfolding the pooled block back into many sites.

Finally, to ensure that the pooled pixels which originally formed spatially contiguous blocks remain adjacent in the one-dimensional MPS chain, we order the sites according to a Z-order (Morton) curve [6] rather than the usual row-major scan (see Figure 1). This space-filling curve preserves locality: pixels that were neighbors in $H \times W$ blocks remain neighbors in the MPS. We gradually reduce the exponent (or increase the resolution) until we recover the full $N = 784$ MPS. This curriculum allows us to reach the full-resolution model much more quickly than naively training on high-resolution data from the outset (see Figure 2 for a comparison).

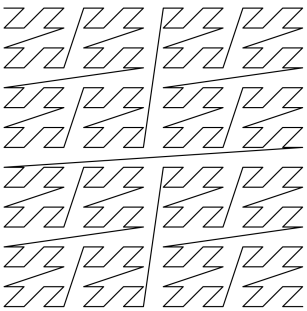

Figure 1. A 16×16 Z-order curve.

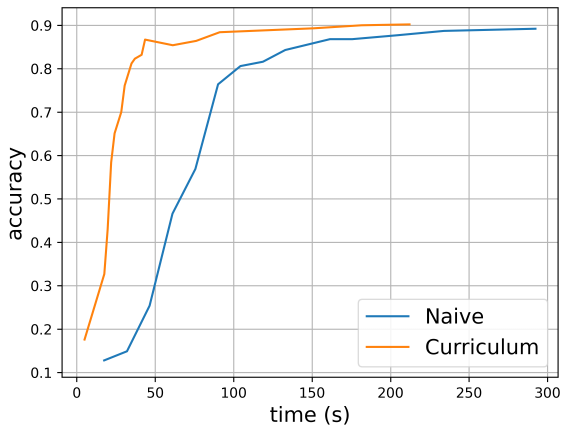

Figure 2. Accuracy vs. time for MPS training.

Ⅲ. Conclusion

In this work, we have introduced a novel curriculum learning framework for MPS models applied to high-dimensional sequence data such as images. By starting from low-dimensional inputs and gradually increasing resolution, our method accelerates convergence and reduces overall training time. The use of a Z-order space-filling curve ensures that spatial locality is preserved throughout the curriculum.

Empirical evaluations on MNIST demonstrate that our curriculum MPS achieves comparable or superior classification accuracy to naive full-resolution training, while requiring fewer contraction operations in the early stages and converging more rapidly. Moreover, our approach preserves parameter efficiency and interpretability inherent to MPS representations. We believe that curriculum learning via exponentiated contractions and locality-preserving orderings provides a paradigm for scalable tensor network training.

ACKNOWLEDGMENT

This work is in part supported by the National Research Foundation of Korea (NRF, RS-2024-00451435 (40%), RS-2024-00413957 (40%)), Institute of Information & communications Technology Planning & Evaluation (IITP, RS-2021-II212068 (20%)) grant funded by the Ministry of Science and ICT (MSIT), Institute of New Media and Communications (INMAC), and the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2025.

REFERENCES

[1] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[2] Perez-Garcia, David, et al. "Matrix product state representations." arXiv preprint quant-ph/0608197 (2006).

[3] Orús, Román. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states." Annals of physics 349 (2014): 117-158.

[4] Stoudenmire, Edwin, and David J. Schwab. "Supervised learning with tensor networks." Advances in neural information processing systems 29 (2016).

[5] Bengio, Yoshua, et al. "Curriculum learning." Proceedings of the 26th annual international conference on machine learning. 2009.

[6] Morton, Guy M. "A computer oriented geodetic data base and a new technique in file sequencing." (1966).