2019 IEEE International Conference on Acoustics, Speech and Signal Processing

Randomized Tensor Ring Decomposition and Its Application to Large-scale Data Reconstruction

Longhao Yuan[1, 2], Chao Li[2], Jianting Cao[1, 2, 4], and Qibin Zhao[2, 3]

[1]Saitama Institute of Technology, [2]RIKEN Center for Advanced Intelligence Project, [3]Guangdong University of Technology, [4]Hangzhou Dianzi University

Contact: longhao.yuan@riken.jp

Background

- Dimensionality reduction is an essential technique for multi-way large-scale data, i.e., tensor.
- Tensor ring decomposition (TRD) has become popular due to its high representation ability and flexibility.
- The existing TRD algorithms suffer from high computational cost when facing large-scale data.
- Random projection (TP) of matrix has been widely applied to solve large-scale problems.
- Tensor random projection (TRP) is a promising tool to solve large-scale tensor problems.

Tensor ring decomposition

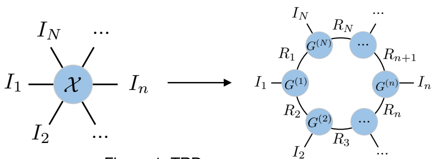

Figure 1. TRD

Decompose tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ in TR-format:

$$\mathcal{X}(i_1, i_2, \ldots, i_N) = \text{Trace}\left\{\prod_{n=1}^{N} \mathbf{G}_n(i_n)\right\}$$

TR core tensor: $\mathcal{G}_n \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}$

TR-rank: $\{R_1, R_2, \ldots, R_{N+1}\}, R_1 = R_{N+1}$

$\mathbf{G}^{(n)}_{i_n}$ is the *i*th slice matrix of the *n*th core tensor.

- **Beyond CP and Tucker**: the curse of dimensionality free, super compressibility
- **Beyond TT format**: enhanced representation ability, permutation flexibility of the latent factors, structure information interpretability
- **Existing problems**: in demand of efficient TRD algorithms

Tensor random projection

We extent matrix random projection to tensor by processing random projection at every mode of the tensor, then a much smaller subspace tensor is obtained which reserves most of the actions of the original tensor. The TRP is simply formulated as follows:

$$\mathcal{X} \approx \mathcal{X} \times_1 \mathbf{Q}_1 \mathbf{Q}_1^T \times_2 \cdots \times_N \mathbf{Q}_N \mathbf{Q}_N^T \approx \mathcal{P} \times_1 \mathbf{Q}_1 \times_2 \cdots \times_N \mathbf{Q}_N$$

Orthogonal projection matrices: $\mathbf{Q}_n \in \mathbb{R}^{I_n \times K_n}$

Projected tensor of smaller size: $\mathcal{P} \in \mathbb{R}^{K_1 \times \cdots \times K_N}$

Randomized TRD

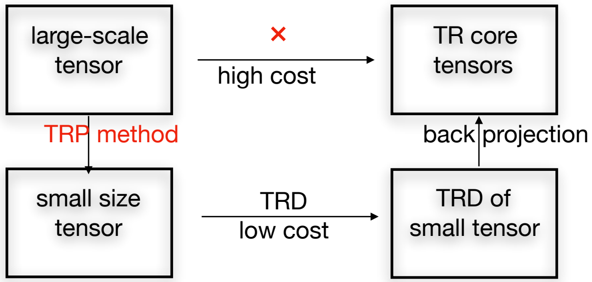

Figure 2. The flowchart of randomized TRD

Given a large-scale tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, the randomized tensor ring decomposition is processed as follows:

1. Find the orthogonal projection matrices by

$$[Q_n, \sim] = QR(\mathbf{X}_{(n)}\mathbf{M})$$

where $\mathbf{M} \in \mathbb{R}^{\prod_{i=1, i \neq n}^{N} I_i \times K_n}$ follows the Gaussian distribution.

2. Process tensor random projection to obtain the projected tensor $\mathcal{P} \in \mathbb{R}^{K_1 \times \cdots \times K_N}$

3. Tensor ring decomposition by TRSVD or TRALS [1] to obtain core tensor $\{\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_N\}$ of $\mathcal{P}$.

4. Back projection to obtain the TRD of the large-scale tensor :

$$\mathcal{G}_n = \mathcal{Z}_n \times_2 \mathbf{Q}_n$$

Remarks:

- The random projection is very effective when the large-scale tensor is low-rank in some modes.
- We can take a balance of the accuracy and computational cost by the choosing proper size of the projection.
- The orthogonal projection matrices can be generated by more efficient methods, to improve the final performance. See more discussions in [2, 3].

Experiments

The experiments show that the proposed algorithms are much faster than traditional algorithms without loss of accuracy, and our algorithms show superior performance in image denoising and deep learning dataset compression compared to the other randomized algorithms.

1. Reconstruction of image of size 5698×4234×3

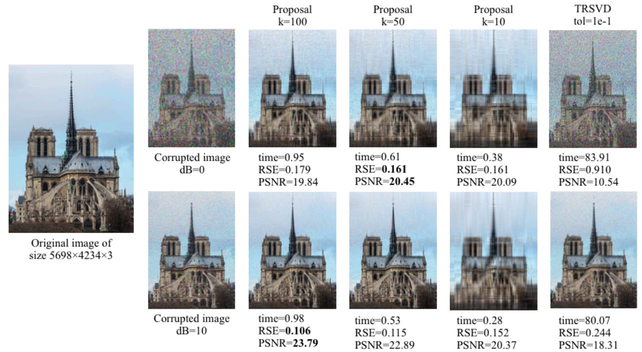

Figure 3. The projection size are selected as {100, 100, 3}, {50, 50, 3}, {10, 10, 3} to see the influence of the projection size. Our method is much faster than TRSVD.

2. Compression of deep learning datasets

Table 1. The compression experiment of two deep learning datasets of size 32×32×3×50000 and 31×31×3×72×100, respectively. The related algorithms are in [4, 5].

	Cifar10						Coil100					
	CR	RSE	time	CR	RSE	time	CR	RSE	time	CR	RSE	time
rTRALS	102.3	0.2185	18.29	767.0	0.3294	17.39	2948.7	0.3331	40.99	1047.3	0.2911	42.61
rTRSVD	42.64	**0.1791**	10.63	42.6	**0.1791**	10.85	175.4	**0.2669**	1.49	175.4	**0.2663**	1.96
TRSGD	102.3	0.4382	1.21e3	767.0	1.00	6.27e2	2948.7	0.4158	482.64	1047.3	0.3536	411.12
rCPALS	99.0	0.2254	11.32	613.6	0.3284	10.86	3084.9	0.3434	2.12	1028.3	0.3001	5.80
rTucker	100.8	0.2146	10.65	509.2	0.3058	4.61	3093.5	0.4241	0.38	1077.4	0.4680	1.98

Conclusion

Based on tensor random projection method, we proposed rTRALS and rTRSVD algorithms, by which, without losing accuracy, the large-scale tensor TRD is much faster and outperforms the compared randomized algorithms in image reconstruction and deep learning dataset compression experiments.

Randomized method is a promising aspect for large-scale data processing. In our future work, we will focus on further improving the performance of decomposition and applying randomized algorithms to very sparse and incomplete tensors.

Reference

[1] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, “Tensor ring decomposition,” arXiv preprint arXiv:1606.05535, 2016.

[2] P. Li, T. j. Hastie, and K.W. Church, “Very sparse random projections,” Proceedings of the 12th ACM SIGKDD, 2006.

[3] E. Bingham, and H. Mannila, “Random projection in dimensionality reduction: applications to image and text data,” Proceedings of the seventh ACM SIGKDD, 2001.

[4] N.B. Erichson, K. Manohar, S. L. Brunton, and J. N. Kutz, “Randomized CP tensor decomposition,” arXiv preprint arXiv:1703.09074, 2017.

[5] G. Zhou, A. Cichocki, and S. Xie, “Decomposition of big tensors with low multilinear rank,” arXiv preprint arXiv:1412.1885, 2014.