ParTI!: a Parallel Tensor Infrastructure for Data Analysis

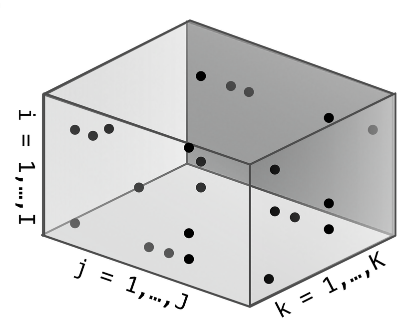
Jiajia Li[1], Yuchen Ma[2], Chenggang Yan[2], Jimeng Sun[1], Richard Vuduc[1]
[1] Georgia Institute of Technology, Georgia, USA
[2] Hangzhou Dianzi University, Hangzhou, China

Background

Tensor decomposition is a set of unsupervised methods to analyze and extract knowledge from tensors, which is widely used in healthcare analytics, image processing, machine learning, and social network analytics. Basic tensor operations are the computational kernels of tensor decomposition algorithms.

Sparse tensors

Many real-world tensors are hyper-sparse and have specific features. To discover useful knowledge, efficient sparse algorithms are critical to performance and scalability.



A 3rd-order sparse tensor

Tensor Operations

Basic tensor operations include element-wise operations, Kronecker product, Khatri-Rao product, sparse tensor times matrix (SpTTM) product, matricized tensor times Khatri-Rao product (MTTKRP), and tensor matricization.

Contributions

- This work distinguishes structured sparse tensors with a small number of dense modes from general sparse tensors by using a new data structure (sCOO) and optimizes SpTTM algorithm based on it.
- This work introduces a large-scale parallel tensor infrastructure (ParTI!) for arbitrary-order sparse tensors on both multicore CPU and GPU platforms.
- The accelerated operations include element-wise tensor operations, SpTTM, and MTTKRP, which are the computational kernels for popular tensor decomposition algorithms (CP and Tucker decompositions).

Data Structures

COO format: general sparse tensors

sCOO format: semi-sparse tensors with one or several dense modes.

i	j	k	val
1	1	1	1
1	1	2	2
2	1	1	3
2	1	2	4
3	3	1	5
3	3	2	6

(a) COO

i	j	val
1	1	1 2
2	1	3 4
3	3	5 6

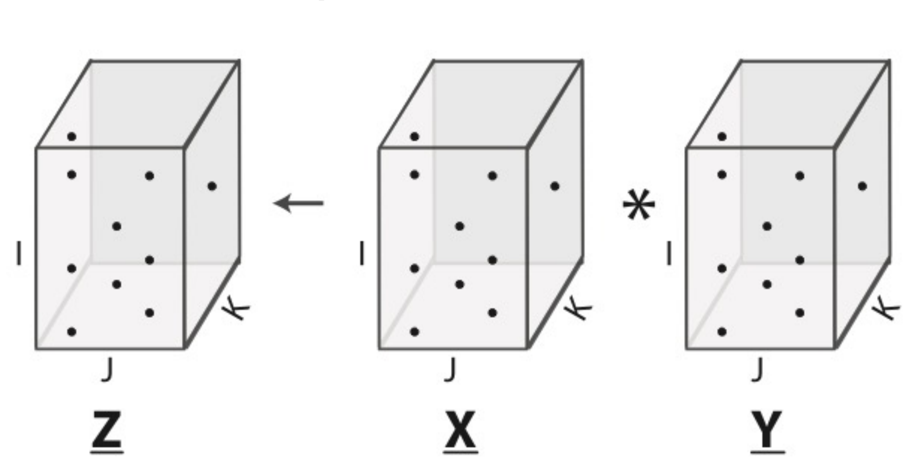
(b) sCOO

sCOO benefits:

- Space-efficient, save at lease k/(N+1) storage for a N^{th}-order semi-sparse tensor with k dense modes.
- Fast contiguous memory access for the dense modes.

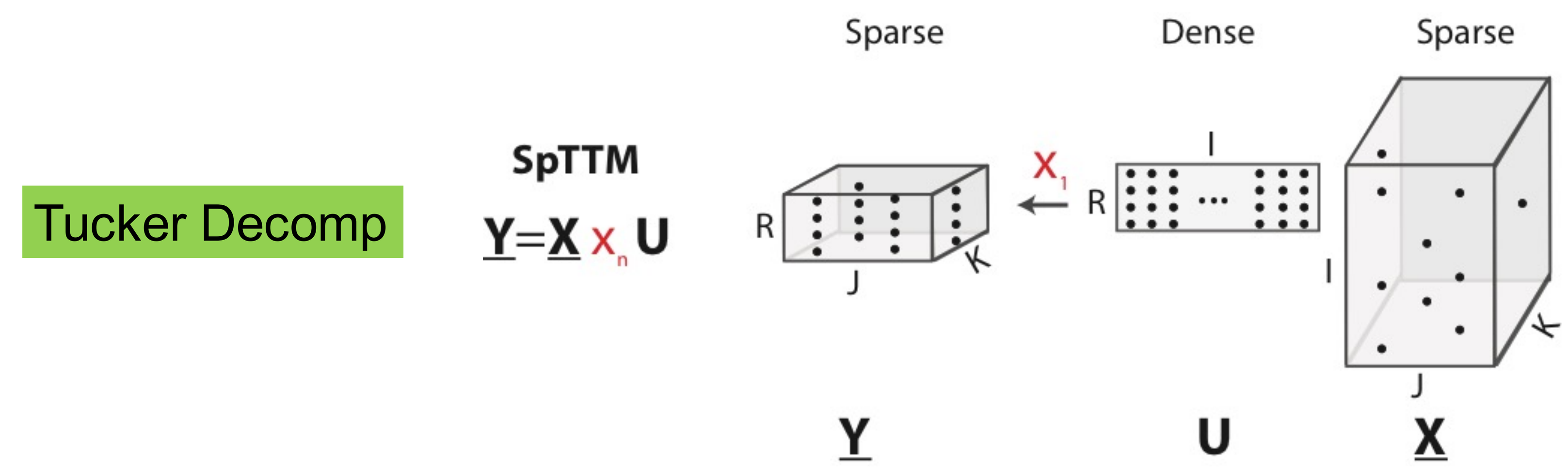
Operations

Hadamard Product (Element-wise Tensor Product)

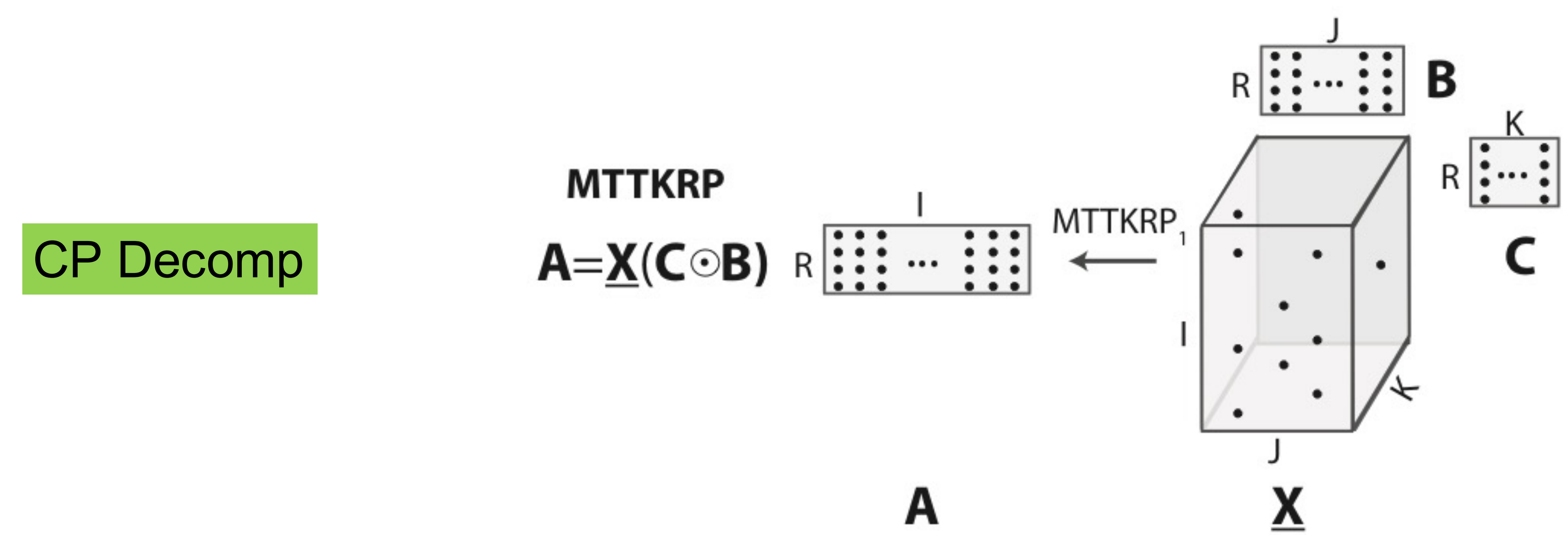


Operations

Sparse Tensor Times Matrix (SpTTM)



Matricized Tensor Times Khatri-Rao Product (MTTKRP)



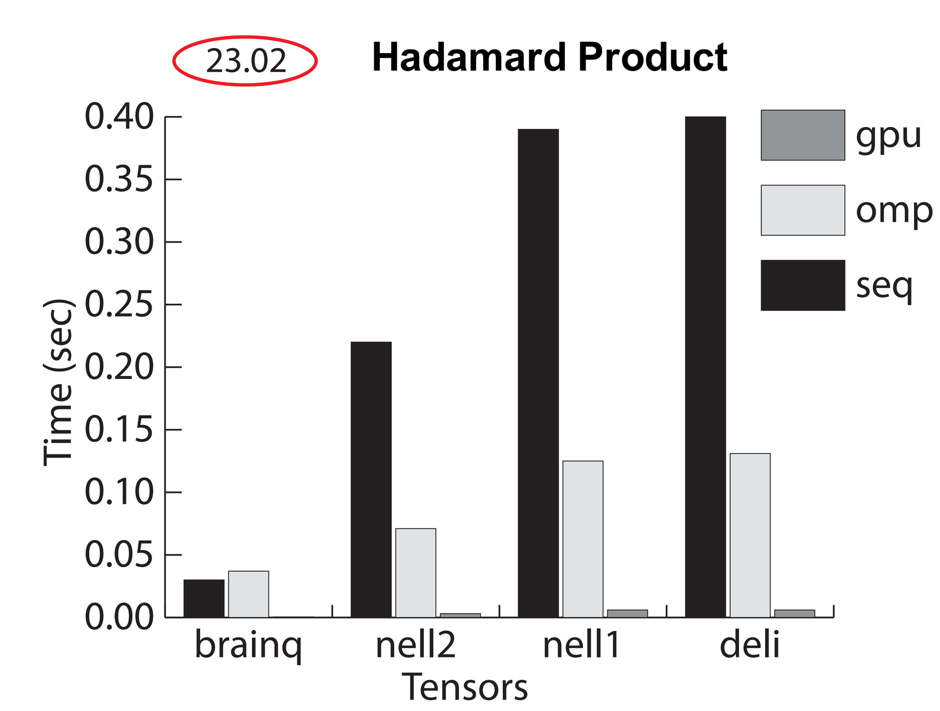
Results

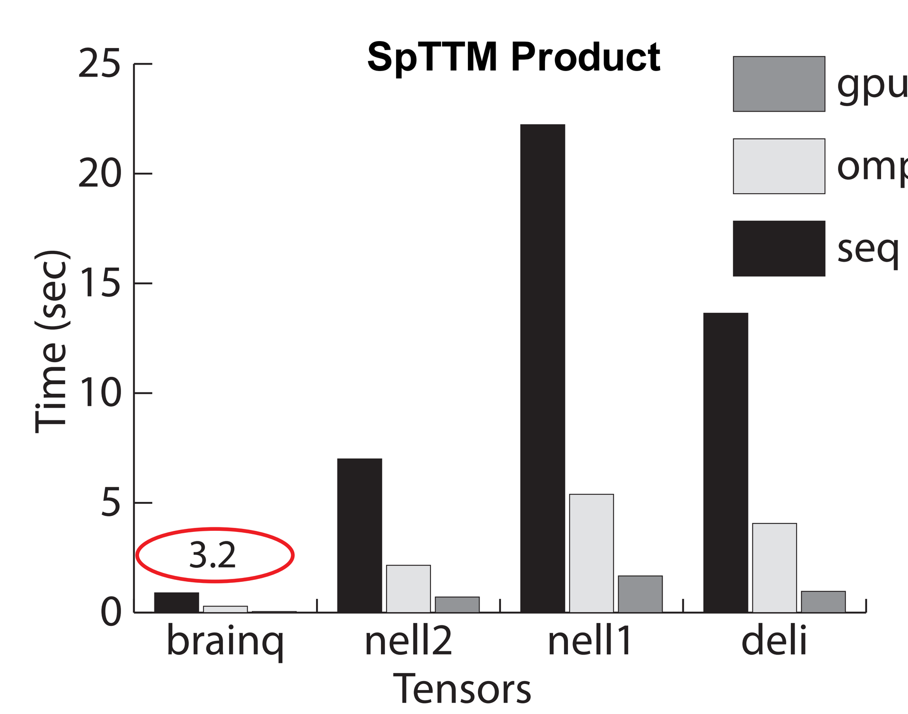
We test our algorithms on Intel Core i7-4770K and NVIDIA Tesla K40c platforms. The sparse tensors are from functional Magnetic Resonance Imaging (fMRI) measurements of brain activity, Never Ending Language Learning (NELL) project, and data crawled from tagging systems.

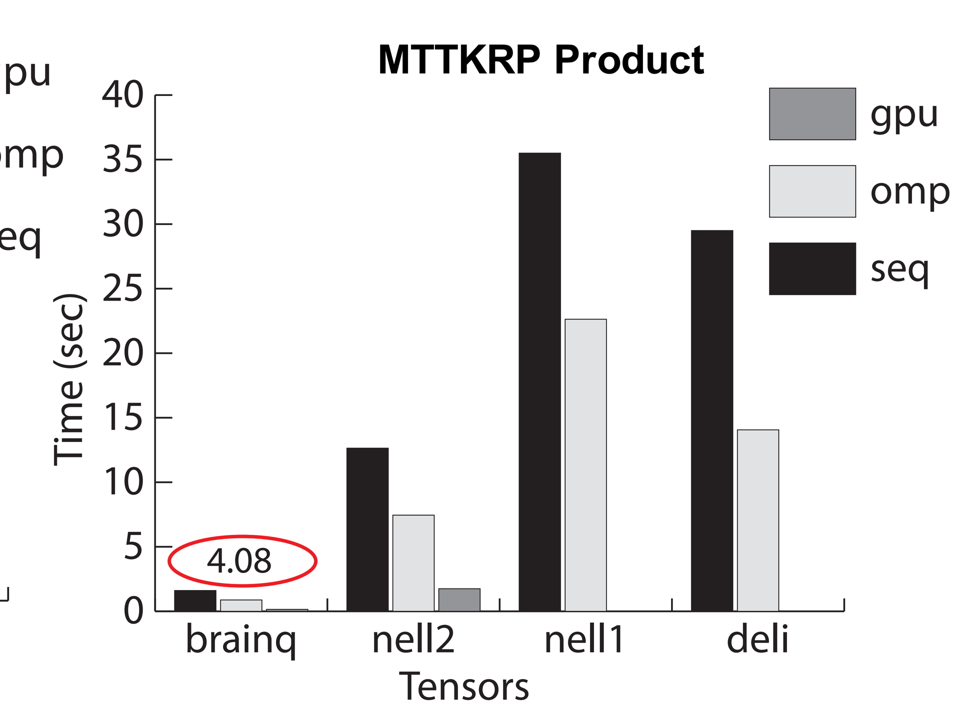
Dataset	Order	Mode sizes	NNZ	Density
brainq	3	$60 \times 70K \times 9$	11M	2.9e-01
nell2	3	$12K \times 9K \times 29K$	77M	1.3e-05
nell1	3	$2.9M \times 2.1M \times 25.5M$	144M	3.1e-13
deli	3	$0.5M \times 17.3M \times 2.5M$	140M	6.1e-12

Sparse tensor description

Matlab Tensor Toolbox







Conclusion

ParTI! provides high-efficient computational tensor operations for sparse tensor decompositions for a single PC with GPUs.

Future, we will integrate more operations and sparse tensor decompositions on GPU.

References

B. W. Bader, T. G. Kolda et al., "Matlab tensor toolbox (Version 2.6)," Available online, February 2015.

M. Baskaran, B. Meister, N. Vasilache, and R. Lethin, "Efficient and scalable computations with sparse tensors," in High Performance Extreme Computing (HPEC), 2012 IEEE Conference on, Sept 2012, pp. 1–6.

Acknowledgement: This project is supported by the U.S. National Science Foundation (NSF) Award Number 1533768, Zhejiang Province Nature Science Foundation of China LR17F030006, and National Nature Science Foundation of China (61671196, 61327902).