

Parallel Hierarchical Clustering using Rank-Two Nonnegative Matrix Factorization

Lawton Manning¹ Grey Ballard¹
Ramakrishnan Kannan² Haesun Park³

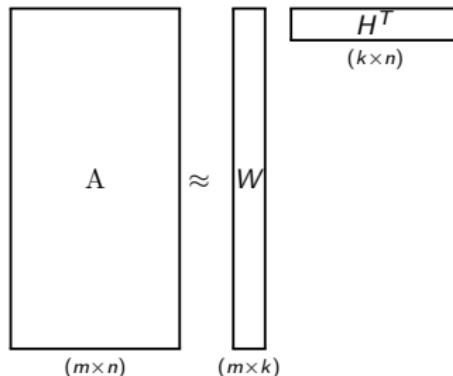
¹Wake Forest University

²Oak Ridge National Laboratory

³Georgia Institute of Technology

27th IEEE International Conference on High Performance
Computing, Data, & Analytics (HiPC 2020)

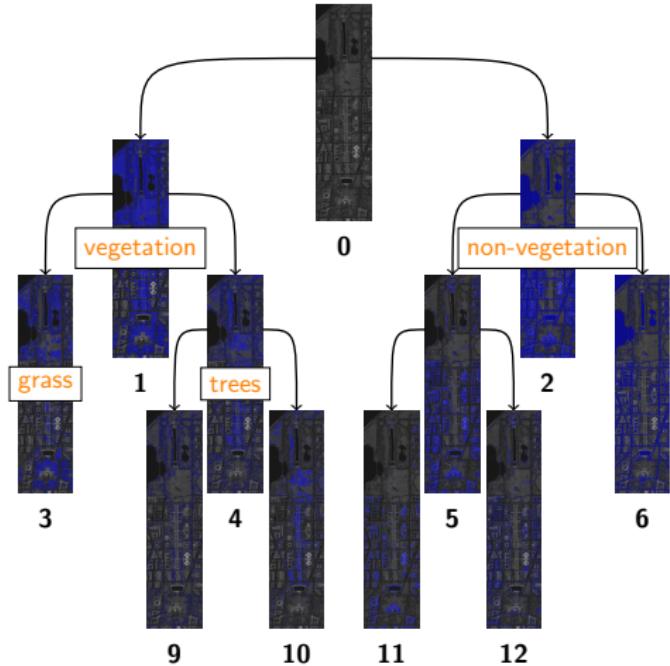
Nonnegative Matrix Factorization (NMF)



- Approximate A (features \times samples) into:
 - W (features \times clusters): feature signatures
 - H (samples \times clusters): cluster membership
- Clustering applications
 - text document topic modeling
 - hyperspectral image segmentation

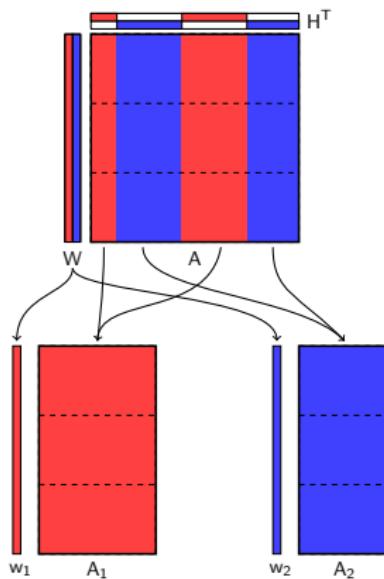
Hierarchical Clustering

- Recursively bipartition data to form binary tree
- Decision choices
 - how to split
 - Rank-2 NMF
 - choosing node to split
 - based on Rank-2 NMF



Splitting with Rank-2 NMF

- Use columns of H^T to bipartition the columns of A
- Assign the cluster columns into two submatrices
- Assign columns of W as feature signatures of clusters



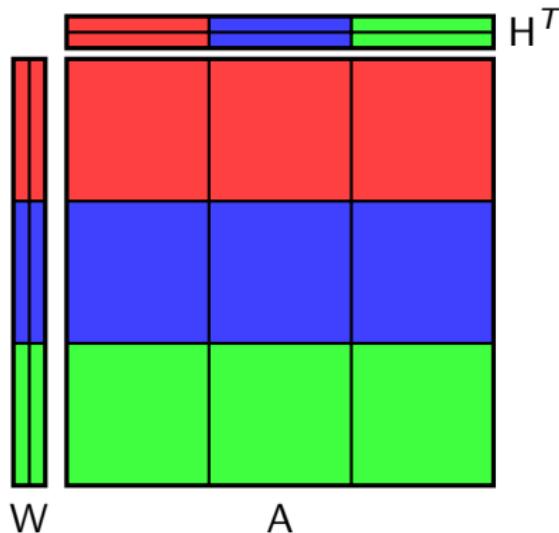
Solving NMF

$$\min_{W,H \geq 0} \|A - WH^T\|_2$$

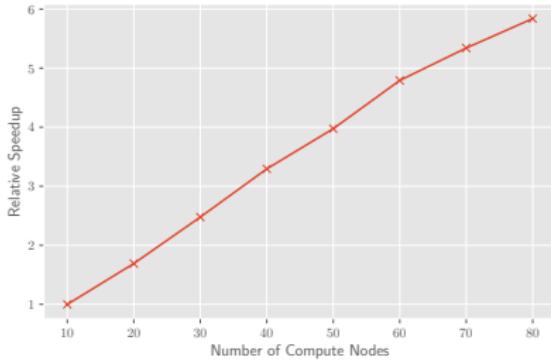
- Alternating Nonnegative Least Squares (ANLS)
 - fix H , solve for W using NLS
 - alternate and repeat
- Rank-2 NLS
 - optimized version of general rank- k algorithm
 - 4 possibilities: ++, +0, 0+, 00
 - solve each and choose optimal

Parallel Rank-2 NMF

- Use row distribution of A, W, and H
 - example distribution across 3 processors shown below
- Computational bottlenecks are matrix multiplications
 - compute $W^T A$ using reduce-scatter for H
 - compute $A H$ using all-gather for W
 - compute $W^T W$ and $H^T H$ using all-reduce

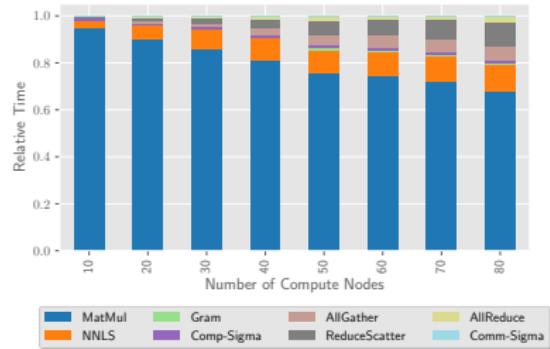


Hierarchical Clustering Scaling and Breakdown



SIIM-ISIC Data

- SIIM-ISIC ($3,145,728 \times 33,126$, 800GB)
 - Society for Imaging Informatics in Medicine - International Skin Imaging Collaboration image classification of melanoma images
- Reasonable scaling that is dominated by MatMul



SIIM-ISIC Data

Conclusions and Future Directions

- Scalability is possible when MatMul dominates, particularly when input data is tall and skinny and tree is balanced
- Possibilities for improving scalability
 - general 2D data distribution
 - parallelization across hierarchical tree
- HierNMF2 as a preprocessing step for Flat NMF