Sparse Local Embeddings for Extreme Multi-label Classification

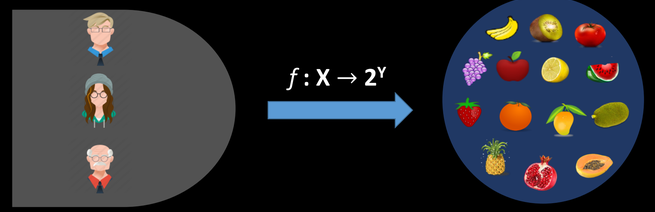
Kush Bhatia† Himanshu Jain§ Purushottam Kar‡ Manik Varma† Prateek Jain†

† Microsoft Research, Bengaluru, India § Indian Institute of Technology Delhi ‡Indian Institute of Technology Kanpur

Microsoft Research

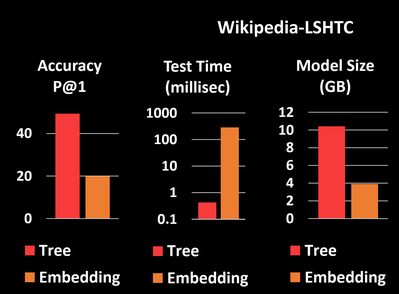
Extreme Multi-Label Learning

- Learning with millions of labels
- New paradigm for ranking and recommendation



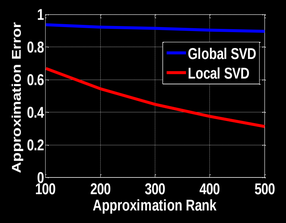
Motivation

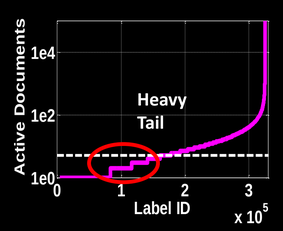
Embedding v/s Tree Techniques:



	Tree	Embedding
Accuracy	✓	✗
Prediction Time	✓	✗
Model Size	✗	✓
Theoretical Guarantees	✗	✓

Limitations of Embedding Methods





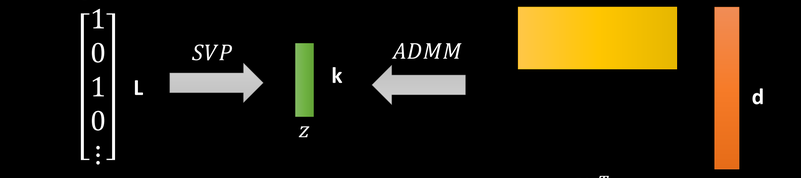
Key Insight:

- k-NN consistently performs very good in heavy tailed distributions.
- Practically infeasible: Test time is O(Nd')

Our Solution : Neighborhood preserving low rank embeddings

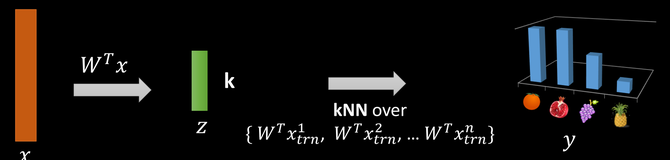
SLEEC : Algorithm

Training:

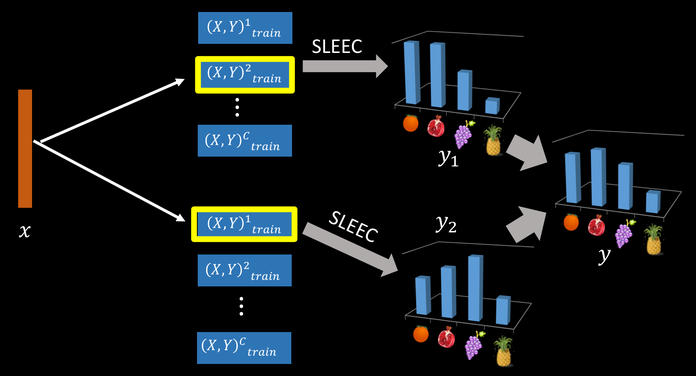


$$\min_{\substack{M \geq 0 \\ rank(M) \leq k}} ||P_\Omega(Y^T Y) - P_\Omega(M)||_F^2 \qquad \min_{W \in R^{d \times k}} ||Z - XW||_F^2 + \lambda ||W||_F^2 + \mu ||XW||_1$$

Prediction:



SLEEC : Scaling up with Multiple Learners



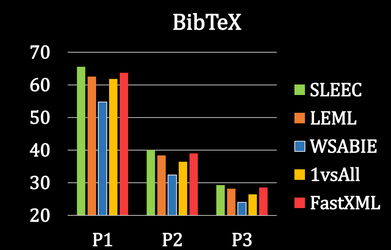
SLEEC : Key Features

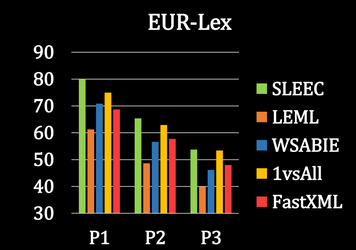
- Can scale efficiently to datasets with a million labels
- Accuracy Improvements of upto
  - 35% over embedding methods
  - 6% over tree methods
- Theoretical Guarantees : Consistency Guarantees for Finite Sample

Results

Dataset Statistics:

Data Set	# of Training Points	# of Test Points	# of Dimensions	# of Labels
EurLEX	17413	1935	5000	3993
BibTeX	4,880	2,515	1,836	159
Wikipedia-LSHTC	1.77M	587k	1.61M	325k





Wikipedia-LSHTC