Discovery of Latent Factors in High-dimensional Data Using Tensor Methods

Furong Huang

University of California, Irvine

Machine Learning Conference 2016 New York City

Machine Learning - Modern Challenges

Big Data Challenging Tasks

Success of Supervised Learning

Image classification

Speech recognition

Text processing

- Computation power growth
- Enormous labeled data

Machine Learning - Modern Challenges

Big Data　　　Challenging Tasks

Real AI requires Unsupervised Learning

Filter bank learning

Feature extraction

Embeddings; Topics

Machine Learning - Modern Challenges

Big Data Challenging Tasks

Real AI requires Unsupervised Learning

Filter bank learning

Feature extraction

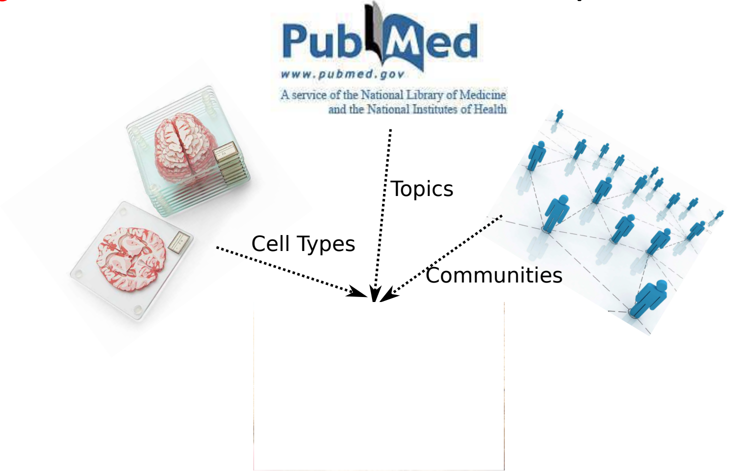
Embeddings; Topics

- Summarize key features in data: Machines vs Humans
- Foundation for successful supervised learning

Unsupervised Learning with Big Data

Information Extraction

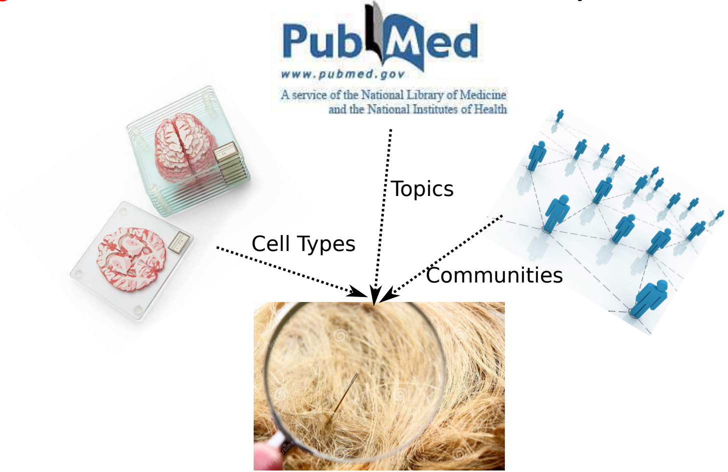
- High dimension **observation** vs Low dimension **representation**



Unsupervised Learning with Big Data

Information Extraction

- High dimension **observation** vs Low dimension **representation**



Finding Needle In the Haystack Is Challenging

Unsupervised Learning with Big Data

Information Extraction

Solution for Unsupervised Learning

$\boxed{\text{A Unified Tensor Decomposition Framework}}$

Automated Categorization of Documents

Mixed topics

SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor — I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five

am's second-leading receiver.

On Jan. 10, 2013, a female student at Florida State spotted the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

In the 21 months since, Florida State officials have said little about how they handled the case, which is no

investigated by the federal Depart

Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.

"Can you share any details on the requesting source?" David Perry, the university's police chief, asked the Tallahassee police. Several hours later, Mr.

Topics

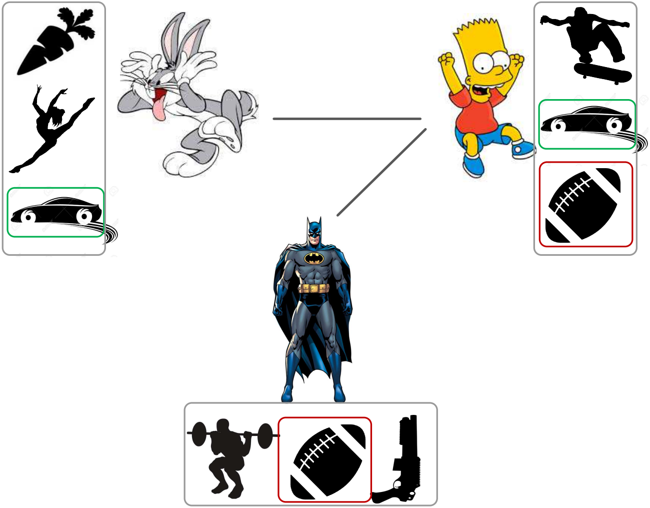
Education

Crime

Sports

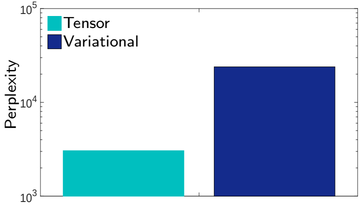
Community Extraction From Connectivity Graph

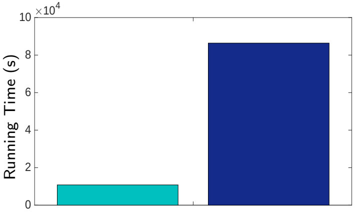
Mixed memberships



Tensor Methods Compared with Variational Inference

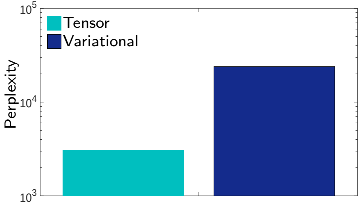
PubMed on Spark: 8 million docs

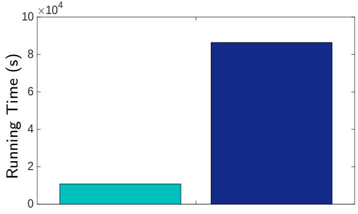




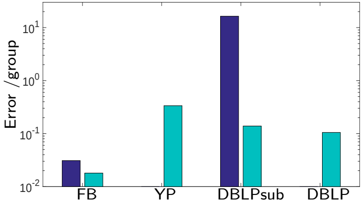
Tensor Methods Compared with Variational Inference

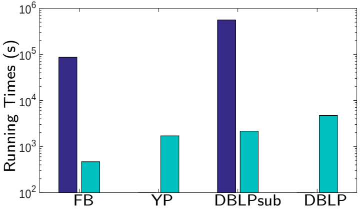
PubMed on Spark: 8 million docs





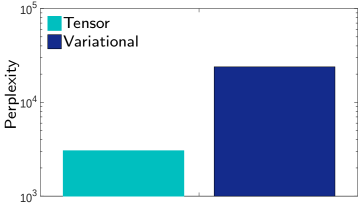
Facebook: $n \sim 20k$ Yelp: $n \sim 40k$ DBLP: $n \sim 1$ million

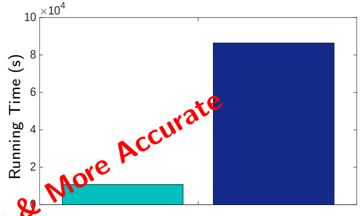




Tensor Methods Compared with Variational Inference

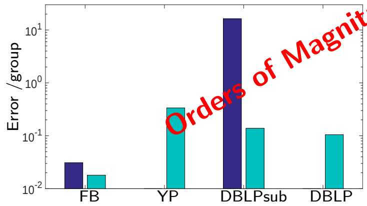
PubMed on Spark: 8 million docs

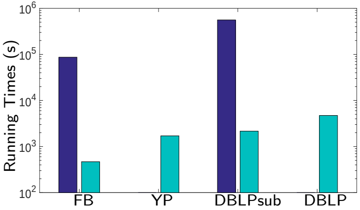




Orders of Magnitude Faster & More Accurate

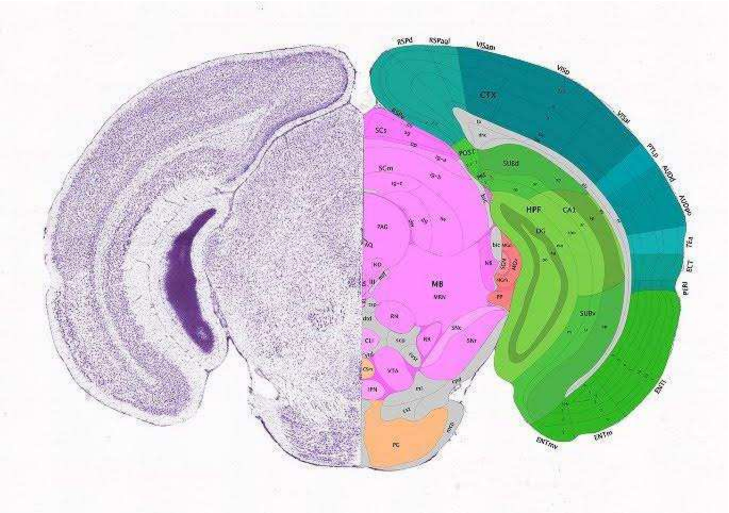
Facebook: $n \sim 20k$ Yelp: $n \sim 40k$ DBLP: $n \sim 1$ million





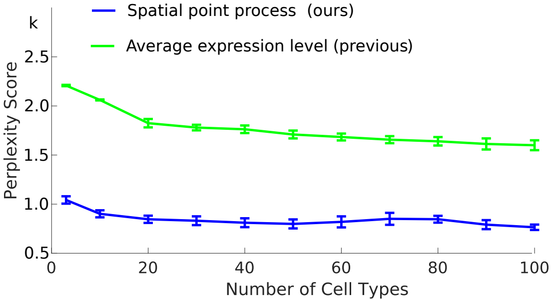
"Online Tensor Methods for Learning Latent Variable Models", F. Huang, U. Niranjan, M. Hakeem, A. Anandkumar, JMLR14.
"Tensor Methods on Apache Spark", by F. Huang, A. Anandkumar, Oct. 2015.

Cataloging Neuronal Cell Types In the Brain



Cataloging Neuronal Cell Types In the Brain

- Our method vs Average expression level [Grange 14']

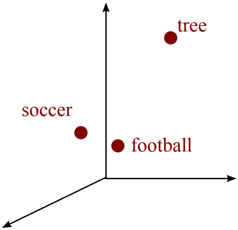


Recovered known cell types

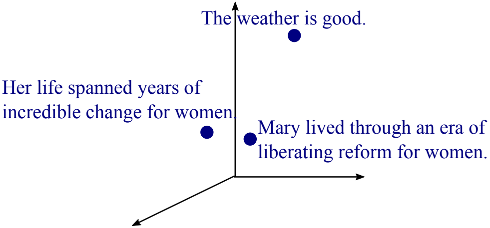
1. *astrocytes*
2. *interneurons*
3. *oligodendrocytes*

" Discovering Neuronal Cell Types and Their Gene Expression Profiles Using a Spatial Point Process Mixture Model " by F. Huang, A. Anandkumar, C. Borgs, J. Chayes, E. Fraenkel, M. Hawrylycz, E. Lein, A. Ingrosso, S. Turaga, NIPS 2015 BigNeuro workshop.

Word Sequence Embedding Extraction

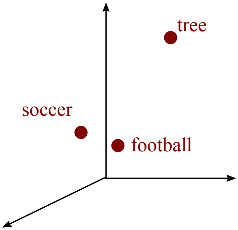


Word Embedding

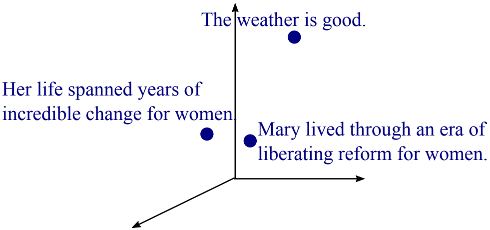


Word Sequence Embedding

Word Sequence Embedding Extraction



Word Embedding



Word Sequence Embedding

Paraphrase Detection

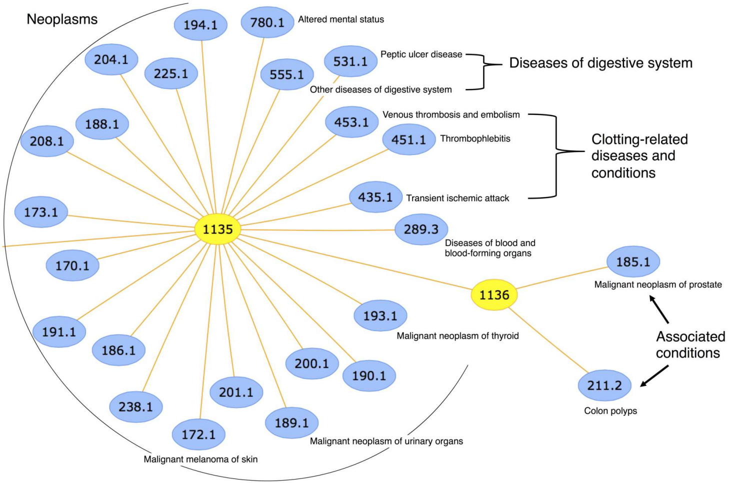
- MSR paraphrase data: 5800 pairs of sentences

Method	Outside Information	F score
Vector Similarity (Baseline)	word similarity	75.3%
Convolutional Tensor (Proposed)	**none**	**80.7%**
Skip-thought (NIPS'15)	train on large corpus	81.9%

"Convolutional Dictionary Learning through Tensor Factorization", by F. Huang, A. Anandkumar, conference and workshop proceeding of JMLR, vol.44, Dec 2015.

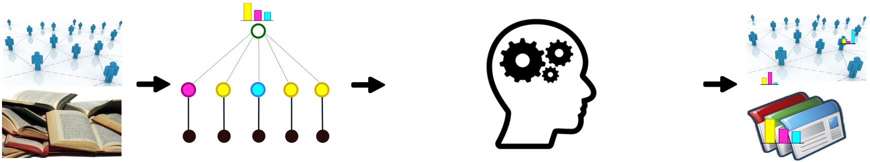
Human Disease Hierarchy Discovery

CMS: 1.6 million patients, 168 million diagnostic events, 11 k diseases.



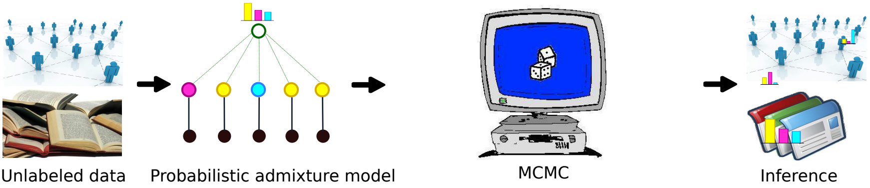
" Scalable Latent TreeModel and its Application to Health Analytics " by F. Huang, N. U.Niranjan, I. Perros, R. Chen, J. Sun, A. Anandkumar, NIPS 2015 MLHC workshop.

Unsupervised Learning via Probabilistic Models



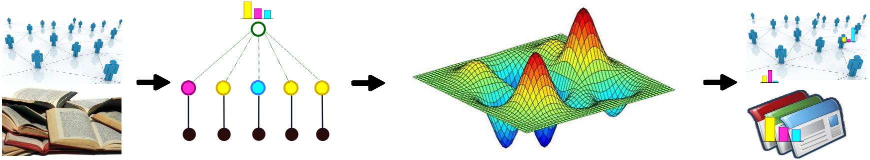
Unlabeled data Probabilistic admixture model Learning Algorithm Inference

Unsupervised Learning via Probabilistic Models



- MCMC: random sampling, slow
  - Exponential mixing time

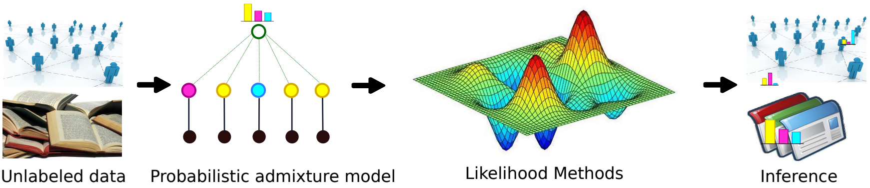
Unsupervised Learning via Probabilistic Models



Unlabeled data Probabilistic admixture model Likelihood Methods Inference

- MCMC: random sampling, slow
  - Exponential mixing time
- Likelihood: non-convex, not scalable
  - Exponential critical points

Unsupervised Learning via Probabilistic Models

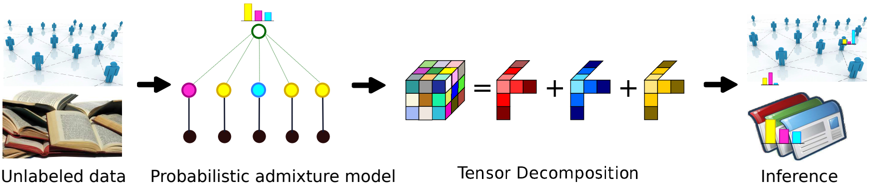


- MCMC: random sampling, slow
  - Exponential mixing time
- Likelihood: non-convex, not scalable
  - Exponential critical points

Solution

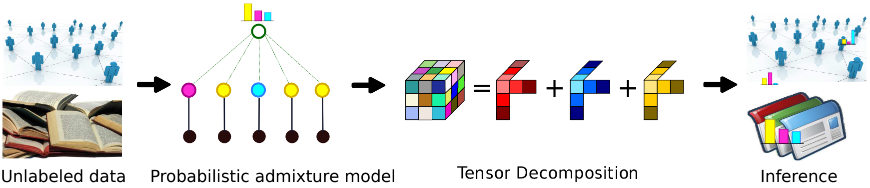
A unified tensor decomposition framework

Unsupervised Learning via Probabilistic Models



$$\boxed{\text{tensor decomposition} \rightarrow \text{correct model}}$$

Unsupervised Learning via Probabilistic Models



$\boxed{\text{tensor decomposition} \rightarrow \text{correct model}}$

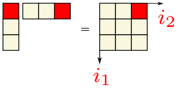
Contributions

- Guaranteed online algorithm with global convergence guarantee
- Highly scalable, highly parallel, random projection
- Tensor library on CPU/GPU/Spark
- Interdisciplinary applications
- Extension to model with group invariance

What is a tensor?

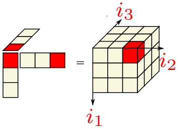
Matrix: Second Order Moments

- M_2: pair-wise relationship.
- $[x \otimes x]_{i_1,i_2} = x_{i_1} x_{i_2} \to [M_2]_{i_1,i_2}$



Tensor: Third Order Moments

- M_3: triple-wise relationship.
- $[x \otimes x \otimes x]_{i_1,i_2,i_3} = x_{i_1} x_{i_2} x_{i_3} \to [M_3]_{i_1,i_2,i_3}$

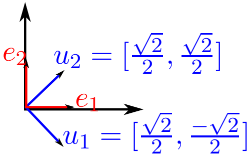


Why are tensors powerful?

Matrix Orthogonal Decomposition

- Not unique without eigenvalue gap

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$

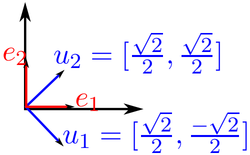


Why are tensors powerful?

Matrix Orthogonal Decomposition

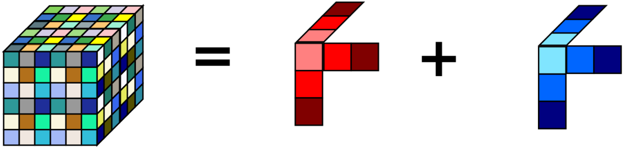
- Not unique without eigenvalue gap

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$



Tensor Orthogonal Decomposition

- Unique: eigenvalue gap not needed

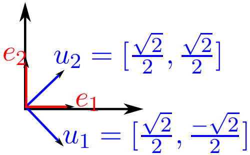


Why are tensors powerful?

Matrix Orthogonal Decomposition

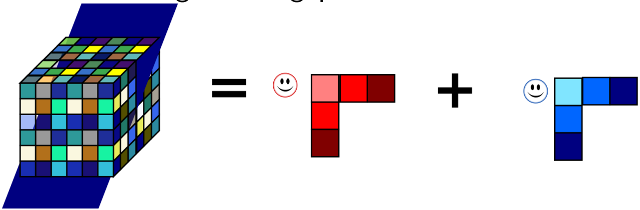
- Not unique without eigenvalue gap

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$



Tensor Orthogonal Decomposition

- Unique: eigenvalue gap not needed
- Slice of tensor has eigenvalue gap

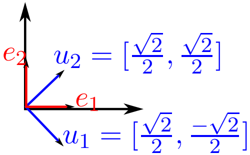


Why are tensors powerful?

Matrix Orthogonal Decomposition

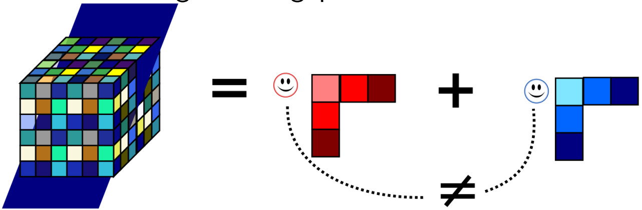
- Not unique without eigenvalue gap

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$



Tensor Orthogonal Decomposition

- Unique: eigenvalue gap not needed
- Slice of tensor has eigenvalue gap



Outline

1. Introduction

2. LDA and Community Models
   - From Data Aggregates to Model Parameters
   - Guaranteed Online Algorithm

3. Conclusion

Outline

1. Introduction
2. LDA and Community Models
   - From Data Aggregates to Model Parameters
   - Guaranteed Online Algorithm
3. Conclusion

Probabilistic Topic Models - LDA

Bag of words

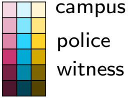
Probabilistic Topic Models - LDA

Bag of words

Probabilistic Topic Models - LDA

Bag of words

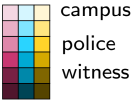
Goal



- Topic-word matrix $\boxed{\mathbb{P}[\text{word} = i | \text{topic} = j]}$

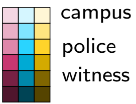
Mixture Form of Moments

Goal: Linearly independent topic-word table

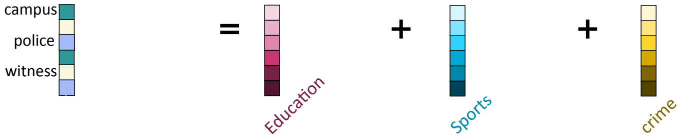


Mixture Form of Moments

Goal: Linearly independent topic-word table

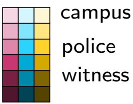


M_1: Occurrence of Words

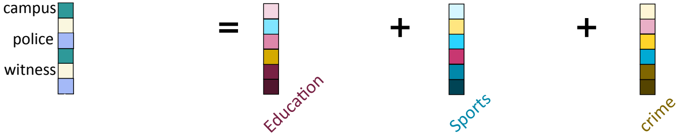


Mixture Form of Moments

Goal: Linearly independent topic-word table



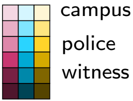
M_1: Occurrence of Words



No unique decomposition of vectors

Mixture Form of Moments

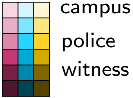
Goal: Linearly independent topic-word table



M_2: Modified Co-occurrence of Word Pairs

Mixture Form of Moments

Goal: Linearly independent topic-word table

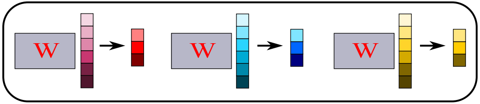


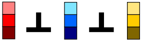
M_2: Modified Co-occurrence of Word Pairs

Matrix decomposition recovers subspace, not actual model

Mixture Form of Moments

Goal: Linearly independent topic-word table

Find a W 

such that 

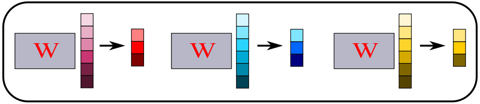
M_2: Modified Co-occurrence of Word Pairs

Many such W's, find one, project data with W

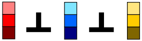
Mixture Form of Moments

Goal: Linearly independent topic-word table

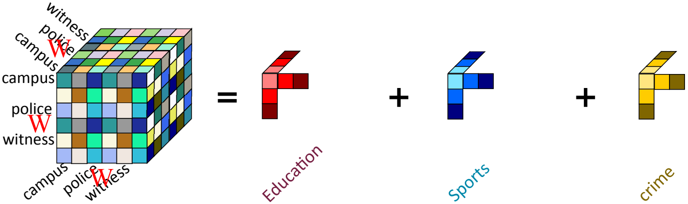
Know a W



such that



M_3: Modified Co-occurrence of Word Triplets

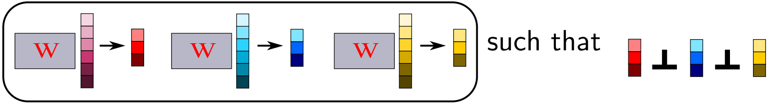


Unique orthogonal tensor decomposition, project result with $W^{\dagger}$

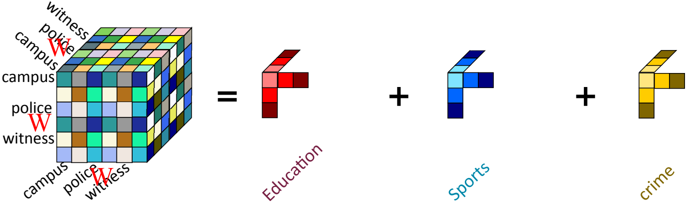
Mixture Form of Moments

Goal: Linearly independent topic-word table

Know a W



M_3: Modified Co-occurrence of Word Triplets

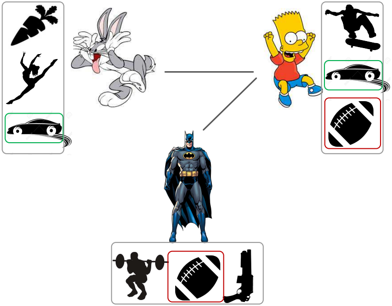


Tensor decomposition uniquely discovers the correct model

Learning Topic Models through Matrix/Tensor Decomposition

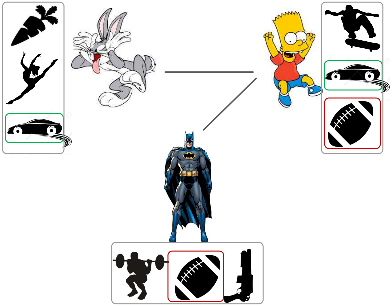
Mixed Membership Community Models

Mixed memberships

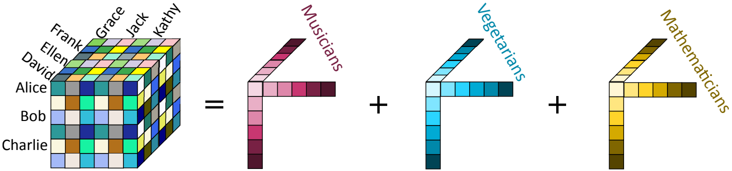


Mixed Membership Community Models

Mixed memberships

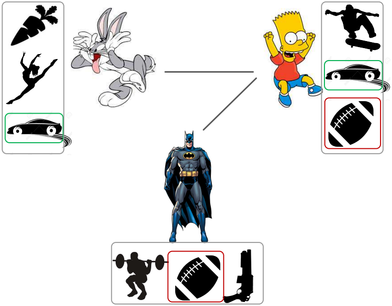


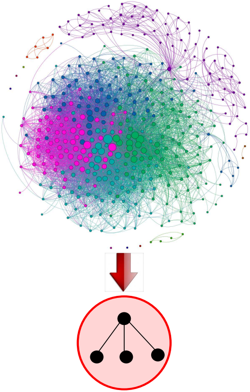
What ensures guaranteed learning?



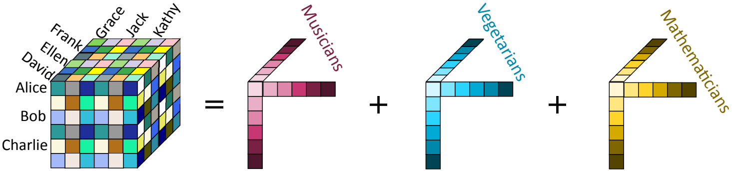
Mixed Membership Community Models

Mixed memberships





What ensures guaranteed learning?



Outline

1. Introduction
2. LDA and Community Models
   - From Data Aggregates to Model Parameters
   - Guaranteed Online Algorithm
3. Conclusion

How to do tensor decomposition?

Model is uniquely identifiable! How to identify?

How to do tensor decomposition?

How to find components? Non-convex optimization problem!

How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

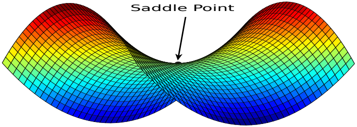
How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

Saddle point: enemy of SGD



- Saddle point has 0 gradient

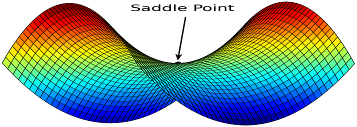
How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

Saddle point: enemy of SGD



- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue

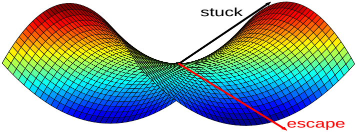
How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

Saddle point: enemy of SGD



- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue
- Negative eigenvalue: direction of escape

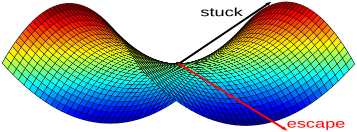
How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

Saddle point: enemy of SGD



- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue
- Negative eigenvalue: direction of escape

Guaranteed Global Converge Online Tensor Decomposition

Theorem: For smooth fn. with non-degenerate saddle points, noisy SGD converges to a local minimum in polynomial steps.

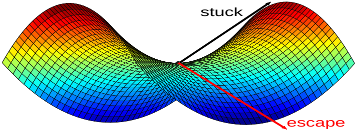
How to do tensor decomposition?

How to find components? Non-convex optimization problem!

Objective Function

Theorem: We propose an objective function with equivalent local optima.

Saddle point: enemy of SGD



- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue
- Negative eigenvalue: direction of escape

Guaranteed Global Converge Online Tensor Decomposition

Theorem: For smooth fn. with non-degenerate saddle points, noisy SGD converges to a local minimum in polynomial steps.

Noise could help!

"Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition",by R. Ge, F. Huang, C. Jin, Y. Yuan, COLT 2015.

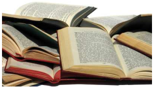
Outline

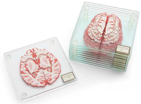
1. Introduction
2. LDA and Community Models
   - From Data Aggregates to Model Parameters
   - Guaranteed Online Algorithm
3. Conclusion

Contributions

Spectral methods reveal hidden structure

- Text/Image processing
- Social networks
- Neuroscience, heathcare ...



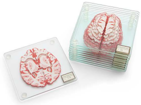


Contributions

Spectral methods reveal hidden structure

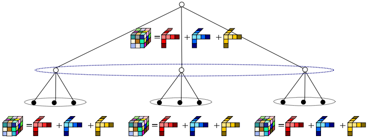
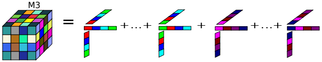
- Text/Image processing
- Social networks
- Neuroscience, heathcare ...

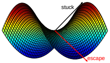




Versatile for latent variable models

- Flat model → hierarchical model
- Sparse coding → convolutional model
- Efficient, convergence guarantee

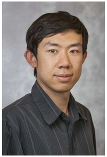





Thank You

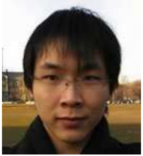
Collaborators

Anima Anandkumar
UC Irvine



Rong Ge
Duke University

Srini Turaga
Janelia Research



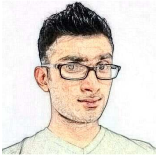
Chi Jin
UC Berkeley

Jennifer Chayes
MSR

Christian Borgs
MSR

Ernest Fraenkel
MIT

Yang Yuan
Cornell U



UN Niranjan
UC Irvine