AFRL-AFOSR-VA-TR-2019-0353

Learning Mixed Membership Community Models - A statistical and a Computational Framework

Animashree Anandkumar
UNIVERSITY OF CALIFORNIA IRVINE

11/06/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
22-11-2019

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
15 May 2015 to 14 May 2019

4. TITLE AND SUBTITLE
Learning Mixed Membership Community Models - A statistical and a Computational Framework

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0221

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Animashree Anandkumar

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF CALIFORNIA IRVINE
5171 CALIFORNIA AVE STE 150
IRVINE, CA 92617 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2019-0353

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
*Developed tensor sketching methods for efficient dimensionality reduction in large datasets
*Analyzed and derived guarantees of tensor sketches and the advantages over previous methods
*Showed gains in deep learning applications using these sketching techniques

15. SUBJECT TERMS
Complex Networks, Hierarchical Bayesian Model, Machine Learning

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UU		NGUYEN, TRISTAN
					19b. TELEPHONE NUMBER *(Include area code)* 703-696-7796

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release.

Final Report for AFOSR Grant FA9550915-1-0221

Anima Anandkumar

August 2019

1 Science

1. What is the mathematical objective of your project? What question are you trying to answer?

   Today we are facing a "data deluge" in almost every domain. Online social networks have seen an explosion in activity and have fundamentally transformed the nature of human interaction. In the biological realm, modern genome sequencers can output data at a rate 400 times faster than the ones a decade ago, and so on. However, although having a transformative potential, the data deluge has not yet been exploited to the fullest extent. Ironically, the data deluge has also resulted in a "data desert". The collected data in many domains are noisy, subsampled, with typically a large number of variables or "unknowns" compared to the number of observations or the "knowns". Such high-dimensionality entails practical principled approaches for learning from ill-posed and ill-behaved data.

   Some of the fundamental questions in high-dimensional learning are: Can we design **scalable** models for efficiently representing and learning high-dimensional data? Here, scalability refers to low **computational** requirements and reduced **sampling** of high-dimensional data. Not all phenomena can be learnt in a scalable manner. Can we characterize the **fundamental limits** on complexity of learning complex phenomena?

2. What are the challenges in doing this? What makes it difficult?

   Learning in high dimensional regime is an ill-posed problem. It is akin to finding a needle in the haystack. In the worst case, learning entails exponential requirements in the amount of samples and computation. Most machine learning tasks require solving non-convex optimization problems. For these problems, as the data dimensions grow, the number of critical points can grow exponentially. Local search methods such as gradient descent can get stuck in one of these critical points. Finding the globally optimal solution is computationally hard in the worst case for non-convex optimization.

3. What is the scientific opportunity that is enabling you to make progress in this difficult area?

   Instead, over the last few years, focus has shifted in characterizing transparent conditions for non-convex problems which are tractable. My research shows that in many instances, these conditions turn out to be mild and natural for machine learning applications. I demonstrate that a range of learning problems can be solved efficiently using tensor methods. The algorithms that I have developed are not only of theoretical interest, but have also been impactful in practice. My methods are scalable to enormous datasets in social networks, document categorization and recommendation systems.

   I have developed a novel class of algorithms based on **tensor factorization of higher order moments of the observed data.** Tensors are higher order generalizations of matrices, and are essential for representing higher order relationships in data. I have characterized how higher order moments (typically third or fourth order) computed from data can be used to learn the parameters of various hidden variable models such as *Gaussian mixtures, latent Dirichlet allocation*, *hidden Markov models*, *network community models*, and so on [1, 2, 3, ?, 4]. These models are relevant in a wide range of applications in domains such as social network analysis, computational biology, document categorization, time series modeling, and so on. I have developed novel algorithms to manipulate and factorize the moment tensors to learn the model parameters. These algorithms retain the computational efficiency of many of the linear algebraic matrix method, and are embarrassingly parallel, which make them ideal for modern machine learning on large-scale data. Thus, I have established that, for the first time, tensor algorithms can accurately learn the parameters of many latent variable models with polynomial computational and sample complexity. The conditions for success are mild, and involve natural non-degeneracy conditions on the model parameters. Tensor algorithms are more than two orders of magnitude faster than previous techniques for learning latent variable models (e.g. variational inference), while also achieving better accuracy of learning [5].

Sketching is a randomized dimensionality-reduction method that aims to preserve relevant information in large-scale datasets. Count sketch is a simple popular sketch which uses a randomized hash function to achieve compression. In [6], we propose a novel extension known as Higher-order Count Sketch (HCS). While count sketch uses a single hash function, HCS uses multiple (smaller) hash functions for sketching. HCS reshapes the input (vector) data into a higher-order tensor and employs a tensor product of the random hash functions to compute the sketch. This results in an exponential saving (with respect to the order of the tensor) in the memory requirements of the hash functions, under certain conditions on the input data. Furthermore, when the input data itself has an underlying structure in the form of various tensor representations such as the Tucker decomposition, we obtain significant advantages. We derive efficient (approximate) computation of various tensor operations such as tensor products and tensor contractions directly on the sketched data. Thus, HCS is the first sketch to fully exploit the multi-dimensional nature of higher-order tensors. We apply HCS to tensorized neural networks where we replace fully connected layers with sketched tensor operations. We achieve nearly state of the art accuracy with significant compression on the image classification benchmark.

As part of this project, [7], we extended the popular LDA model to a general class of topic models, and provided guaranteed spectral methods for learning. We overcome the limitation of LDA to incorporate arbitrary topic correlations, by assuming that the hidden topic proportions are drawn from a flexible class of Normalized Infinitely Divisible (NID) distributions. NID distributions are generated by normalizing a family of independent Infinitely Divisible (ID) random variables. The Dirichlet distribution is a special case obtained by normalizing a set of Gamma random variables. We prove that this flexible topic model class can be learnt via spectral methods using only moments up to the third order, with (low order) polynomial sample and computational complexity. The proof is based on a key new technique derived here that allows us to diagonalize the moments of the NID distribution through an efficient procedure that requires evaluating only univariate integrals, despite the fact that we are handling high dimensional multivariate moments. In order to assess the performance of our proposed Latent NID topic model, we use two real datasets of articles collected from New York Times and Pubmed. Our experiments yield improved perplexity on both datasets compared with the baseline.

In another paper [8], as part of this project, we develop efficient extended BLAS kernels for tensor contractions. Tensor contractions constitute a key computational ingredient of numerical multi-linear algebra. However, as the order and dimension of tensors grow, the time and space complexities of tensor-based computations grow quickly. Existing approaches for tensor contractions typically involves explicit copy and transpose operations. In this paper, we propose and evaluate a new BLAS-like primitive STRIDEDBATCHEDGEMM that is capable of performing a wide range of tensor contractions on CPU and GPU efficiently. Through systematic benchmarking, we demonstrate the advantages of our approach over conventional approaches. Concretely, we implement the Tucker decomposition and show that using our kernels yields 100x speedup as compared to the implementation using existing state-of-the-art libraries.

2 Transitions

Describe anything from this project that you transitioned to anybody else. a. Who did you give it to, and what is their organization? b. What did you give them? Code, papers, algorithms etc. c. What eventual application might this enable? d. What was your scientific accomplishment that enabled this?

Answer: We have released many open source software packages based on the tensor algorithms developed as part of this project. For example, we have a Spark package for tensor decomposition at `https://github.com/FurongHuang/SpectralLDA-TensorSpark`. We have also contributed to Tensorly package `https://github.com/tensorly` and actively developing it.

We have also released code for tensor sketching:

`https://github.com/shiyangdaisy23/Higher-order-CountSketch`

We are also collaborating with a number of companies such as NVidia to develop more open source software, based on tensor algorithms, on a variety of platforms such as GPUs.

References

[1] A. Anandkumar, D. Hsu, and S.M. Kakade. A Method of Moments for Mixture Models and Hidden Markov Models. In *Proc. of Conf. on Learning Theory*, June 2012.

[2] A. Anandkumar, D. P. Foster, D. Hsu, S. M. Kakade, and Y. K. Liu. A Spectral Algorithm for Latent Dirichlet Allocation. In *Proc. of Neural Information Processing (NIPS)*, Dec. 2012.

[3] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

[4] A. Anandkumar, D. Hsu, and A. Javanmard S. M. Kakade. Learning Topic Models and Latent Bayesian Networks Under Expansion Constraints. *Preprint. ArXiv:1209.5350*, Sept. 2012.

[5] F. Huang, U.N. Niranjan, M. Hakeem, and A. Anandkumar. Fast Detection of Overlapping Communities via Online Tensor Methods. *ArXiv 1309.0787*, Sept. 2013.

[6] Yang Shi and Anima Anandkumar. Multi-dimensional tensor sketch. In *Proc. of KDD workshop*, 2019.

[7] Forough Arabshahi and Animashree Anandkumar. Spectral methods for correlated topic models. In *Proc. of AISTATS*, 2017.

[8] Yang Shi, UN Niranjan, Animashree Anandkumar, and Cris Cecka. Tensor contractions with extended blas kernels on cpu and gpu. In *High Performance Computing (HiPC), 2016 IEEE 23rd International Conference on*, pages 193–202. IEEE, 2016.