Kiran Vodrahalli

Github || (650) 490-0526 || knv@princeton.edu || LinkedIn

Summary

I am interested in opportunities involving theoretical and applied research work in algorithms, statistics, optimization, and related fields. I am particularly interested in natural language and neuroscience applications.

Education

Princeton University, September 2016 - June 2017 (expected)
M.S.E. Computer Science

Princeton University, September 2012 - June 2016 (expected)
A.B. Mathematics, Certificates in Statistics/Machine Learning; Computer Science

Relevant Coursework: graduate algorithms and machine learning, statistical learning and nonparametric estimation, neural networks, learning theory, online convex optimization, semidefinite and conic programming, information theory, coding theory and random graphs, linear algebra, probabilistic modeling, natural language processing, functional programming, programming languages, systems programming, game theory, graph theory, point-set and algebraic topology, complex analysis, group theory, Galois theory, mathematical statistical mechanics, real analysis, metric geometry, neuroscience, cognitive modeling

Interests

Theoretical Computer Science
Randomized/Approximation Algorithms, Representation Learning, Statistics,
Online Learning, (Nonconvex) Optimization, Probabilistic Models, Information Theory

Applied Machine Learning
Natural Language Understanding, Computational Neuroscience, Computer Vision

Research Projects

Low-dimensional Representations of Semantic Drift in Thought-Space (with Sanjeev Arora and Ken Norman, Fall 2015—)
My senior thesis, advised by Prof. Arora and Prof. Norman (from Princeton Neuroscience Institute), aims to build a model for extracting low-dimensional representations of the semantic content of an fMRI signal recorded during the presentation of stories in various formats. I am developing novel models and accompanying theoretical analysis to represent the context of a story as it **changes over time**, and verifying my representations with predictive zero-shot brain decoding tasks.

Sparse, Low-dimensional and Multimodal Representations of Time Series for Mind-Reading (for COS 513, Fall 2015—Spring 2016)

This work is joint with Lydia Liu (Princeton ORFE '17) and Niranjani Prasad (Grad Student, Princeton CS Department). We investigated the application of sparse canonical correlation analysis (sCCA) as a tool for creating low-dimensional combined representations of EEG/MEG and fMRI brain data. We used two experiments to demonstrate that our low-dimensional representation retained useful information by testing on two datasets: One was a paired EEG-fMRI time series oddball response dataset and the other was a paired MEG-fMRI time series dataset of subjects looking at various types of objects ([Cichy et. al, 2014]). In both instances we outperformed other traditional methods of low-dimensional representation, including PCA and ICA. We submitted our work as a project for COS 513: Foundations of Probabilistic Modeling, taught by Prof. Barbara Engelhardt, and are now in the process of submitting this paper for publication in a peer-reviewed journal.

Learning Shifting Communities Online in the Adversarial Block Model (for APC 529, Fall 2015)

For the final project in APC 529: Coding Theory and Random Graphs, taught by Professor Emmanuel Abbé, I analyzed the Stochastic Block Model (SBM) from the perpsective of online optimization, making use of recent results in the online learning of eigenvectors and the exact recovery setting of the SBM to build a framework for learning communities as they change over time with guaranteed regret bounds.

Solving Word Analogies with Convex Optimization (with Elad Hazan, Spring 2015—)

My final project for COS 511: Theoretical Machine Learning investigated convex loss functions for learning word vectors to solve word analogy problems. Word analogies are of the form king:man :: queen:woman. Given three of the four words, the task is to correctly identify the fourth. Traditionally, this problem is approached in the unsupervised setting and texts are used to learn which words are most relevant. Word vectors, word representations in high-dimensional real space, are often used (particularly in the past few years) as a solution to the analogy problem via dot-product queries, an approach which has recently been validated by [Arora et al (2015)]. I formulated a convex loss with which to train word vectors that in principle keeps the spirit of the dot product query intuition, implemented AdaGrad [Duchi et al (2011)], and trained on word pairs. This project is a work in progress, as I am continuing to validate experimental results.

Comparing Hebbian Semantic Vectors Across Language (for NEU 330, Spring 2015)

My final project for NEU 330, Connectionist Models, focused on building Hebbian neural network word vector models for parallel corpora, with the purpose of evaluating the word vectors based on how similarly the word vectors for translation pairs behaved in their respective corpora. The principle I held throughout the project was simply that changing language should essentially not effect the representation of a word/concept in a high-dimensional vector space. I both proposed methods of evaluation and made use of previously used metrics to evaluate the 9 models considered. The texts used to form the word vectors were Harry Potter and The Philosopher's Stone and its French counterpart. This project won a prize for being the best in the class.

Noun Compounds in Semantic Quad-Space (with Christiane Fellbaum, Fall 2014)
My junior independent work with Dr. Christiane Fellbaum aimed to build a model for analyzing the similarity between noun compounds, which consist of a modifier noun and a head noun, like "life force." Accurate parsing can greatly improve question answering systems for various knowledge bases. For example, medical QA systems must correctly parse noun compounds like "human colon cancer line" to answer questions accurately. I looked at several approaches to analyzing the similarity of noun compounds and built a vector space model of noun compounds, inspired by the papers [Turney 2013] and [Fyshe et al 2013]. I extended Turney's dual-space model to a quad space model and ran it on two large corpora, COCA and GloWbE. I then evaluated the results by comparing to a ground truth provided by Amazon Mechanical Turk workers. My model achieved 88% accuracy on predicting human judgement of noun compound similarity, improving upon another paper's model. In March 2015, my work won a **school-wide award** as a top project out of 100+ nominations by faculty and other students. The 25-under-25 award recognizes particularly good layman-friendly presentations of innovative research.

Estimating Trending Twitter Topics with Count-Min Sketch (for COS 521, Fall 2014)
My final project for COS 521: Advanced Algorithms was joint with Evan Miller (Princeton COS '16) and Albert Lee (Princeton COS '16). We attempted to solve the following problem: Given a time series of Twitter data, can we infer current trending topics on Twitter while appropriately discounting past tweets using a sketch-based approach? We tweaked the Hokusai data structure [Matusevych et al 2012] and implemented our modifications, then ran experiments on Twitter data.

Characterizing Intellectual Interests with SVM (with Sam Wang, Fall 2013—)
The goal of the project is to investigate intellectual interest as a potential phenotypic marker for autism. In order to study whether this hypothesis was plausible, we had survey responses from two groups of people. The Simons Simplex Collection (SSC) dataset is a repository of genetic samples from families where one child is affected with an autism spectrum disorder. We had survey responses from simplex members, the parents of autistic children. The other responses were obtained by polling readers of Professor Wang's political blog. My role in this project was to create a classifier which given a survey response could output a score indicating certainty that the survey respondant had a particular intellectual interest; for instance, the humanities. Currently, we are preparing to submit a paper on our results: The classifier I eventually trained had 94% accuracy for determining intellectual interest, making the survey-classifier pair potentially useful as a tool. Further work is continuing on assessing the extent to which the classifier output can be used as an autism phenotype.

See kiranvodrahalli.github.io/projects/ for more details on these projects.

Talks

Princeton University
The Representation of Language in the Brain (10/2015, Alg-ML Reading Group)
A Brief Survey on Expander Graphs (04/2015, Junior Seminar organized by Zeev Dvir)
A Survey on Image Captioning with Deep Learning (02/2015, COS 598B)
A Survey on Deep Learning for NLP (02/2015, COS 598B)

The full talks are available at kiranvodrahalli.github.io/projects/.

Teaching

Princeton University

Seminar Leader, NLP-ML Reading Group (Spring 2014 — Present)
Grader, COS 226 (Spring 2014)
COS Lab TA (Fall 2013 — Spring 2015)

Mountain View Library, CA

Math tutor, grades 6 – 12. (Fall 2011 — Spring 2012)

Industry Experience

Palantir Technologies, IQE Intern (Summer 2015)

Worked on adding support for distributed systems frameworks for machine learning pipelines with Spark, YARN, and HDFS.

Intel Corporation, PerC Intern (June 2011 — August 2012, Summer 2013)

Worked on basic computer vision algorithms for depth-cameras analagous to Microsoft Kinect, 3D image capture, basic natural language processing, speech recognition evaluation. Investigated the feasibility of using brain-computer interfaces. Made a few gesture-based demos as well.

Skills

Programming Languages

Python, LaTeX, C, OCaml, Haskell, C++, BASH, Mathematica, MATLAB, Java

Distributed Systems

Some basic experience with: Spark, YARN, HDFS