

# KUSHAL ARORA

Email: [kushal18@gmail.com](mailto:kushal18@gmail.com) Phone: +352-871-5169

Github: <https://github.com/kushalarora>

## **EDUCATION:**

**Master of Science, Computer Engineering**

*University of Florida, Gainesville, FL*

**December 2015**

CGPA: 3.74

### **Courses Taken:**

*Maths for Intelligent Systems, Machine Learning, Advanced Machine Learning, Cloud Computing and Storage, Advance Data Structures, Analysis of Algorithms, Computer Architecture, Distributed Operating System*

### **Master's Thesis:**

Compositional Language Modeling ([pdf](#), [code](#))

**B. Tech, Electronics and Communication Engineering**

*Motilal Nehru National Institute of Technology, Allahabad*

**July 2006-May 2010**

## **PUBLICATIONS:**

Grover, S.; Arora, K.; Mitra, S.K., Text Extraction from Document Images Using Edge Information, Proceedings of India Conference of IEEE INDICON'09, Published by IEEE.

## **PREPRINTS:**

Arora, K.; Rangarajan, A.; Compositional Language Model ([pdf](#), [code](#))

Arora, K.; Rangarajan, A.; Contrastive Entropy: A new evaluation metric for sentence level language models ([pdf](#), [code](#))

## **SELECTED PROJECTS:**

**Compositional Language Model**([pdf](#), [code](#))

Traditional language models treat language as a linear chain on words. In my Master's thesis I challenged this assumption and proposed a model that uses underlying compositional structure for modeling language. This is done by marginalizing the joint probability of sentence and composition tree. Composition trees were generated using PCFGs and marginalization was carried out using the Inside algorithm. Conditional probability given the structure was modeled using the distributional representation similar to neural network based models. We report more than 100% improvement in Contrastive Entropy over RNNLM on a toy data set.

**Contrastive Entropy: A new metric for evaluating sentence level language models**([pdf](#))

Perplexity is an unsuitable metric for sentence level language model due to its reliance on exact probabilities. Other extrinsic metrics like WER are costly to calculate and need extensive setup. As part of my thesis, I also proposed a new discriminative metric to evaluate sentence level language models. The intuition here is to capture the model's ability to differentiate between a good sentence and a bad sentence, one that cannot be generated by the language source.

**Sentence Level Recurrent Neural Network**([pdf](#), [code](#))

To prove the efficacy of our new metric, I reformulated RNN as a sentence level compositional model trained discriminatively to differentiate a sentence from the training set and a slightly distorted version of it. On our new metric we reported considerable gains over RNN when this model is trained at lower level of distortion(10%).

# KUSHAL ARORA

Email: [kushal18@gmail.com](mailto:kushal18@gmail.com) Phone: +352-871-5169

Github: <https://github.com/kushalarora>

## **Comparative evaluation of Manifold Learning Algorithms** ([code](#))

Implemented the state of the art dimensionality reduction algorithms in python using scipy and compared them on four data sets, namely RaceSpace, Digits, Faces and Swiss Roll. The project was an individual effort and done as a course project for Advanced Machine Learning class.

*Algorithms implemented:* Local Linear Embedding, ISOMap, Laplacian Eigenmaps, Hessian LLE, Local Target Space Analysis, Stochastic Neighborhood Embedding

## **Comparative evaluation of Supervised Learning Algorithms** ([code](#))

Built a generic framework to run a list of Supervised Learning Algorithms in Python using Scikit-Learn and Theano. The framework was used to do a comparative study on following data sets: Wisconsin Breast Cancer, Iris, Higgs, OCR and Hand Writing Recognition across a range of supervised learning algorithms. This project was done in a team of three for Machine Learning class.

*Algorithms evaluated were:* Multi Layer Preceptron, Stacked Auto Encoders, Deep Belief Network, Support Vector Machine, Random Forest, Decision Tree, AdaBoost Decision Tree.

## **Ontology Alignment for Knowledge Bases** ([code](#))

Implemented and evaluated PARIS, an ontology alignment technique that uses web text based interlingua for aligning relations and entities. Ontologies for Freebase, NELL and Yago were mapped to each other using label propagation algorithm. This project was done as a independent study in Data Science Lab under Dr Daisy Wang and was a part of larger objective to build a master KB for the lab.

## **PROFESSIONAL EXPERIENCE:**

### **Amazon**

**September 2015 - present**

*Software Engineering Intern, Alexa Entertainment*

Worked on launching Spotify on Alexa and supporting music on new devices Echo Dot and Tap.

### **Amazon**

**May 2014 -August 2014**

*Software Engineering Intern, Transactional Risk Management Services*

Analyzed counterfeit spike problem for high volume items on third party marketplace. Also designed a generic framework that flags and block sales of dubious products based on a certain criteria.

### **Chatimity**

**Sept 2011-June 2013**

*Software Engineer*

First employee at Chatimity. Worked with two founders to build the complete technology stack from scratch. Helped build a system that handles 30,000 uniques and 11.5 million messages per day (upto June 2013). Designed and launched central product features including topic pages, search, recommendations, notifications and image sharing.

### **ST-Ericsson**

**Aug 2010-Sept 2011**

*System Software Engineer, Multimedia Audio Team*

Developed OpenMaxIL layer components for Audio 3D Mixer and AAC Encoder and implemented features like http streaming, buffering and seek features at framework level.

## **TECHNICAL SKILLS:**

**Languages:** C, C++, Java, Python, Javascript, CSS, MySql, Matlab, Latex

**Tools:** Git, GDB, MongoDB, Hadoop, Makefiles, Android SDK, Solr, Tornado Web Server, Theano