CURRICULUM VITAE

Siddharth Sachdeva

http://siddsach.github.io

December 2017

FOCUS

Machine Learning and Natural Language Processing (code at Github)

RESEARCH STATEMENT

I'm interested in human-centered machine learning, specifically they applications and connections of natural language processing to human computer interaction and computational social science. Specifically, I'm passionate about

- Interactive interfaces that make large-scale data analysis accessible to non-technical people
- Deep transfer and multi-task Learning to improve generalization in data-sparse contexts
- Methods to improve model interpretability and mitigate bias.

EDUCATION

University of Chicago — Chicago, IL

B.A. in Statistics, Minor in Computer Science — Expected, June 2018

GPA**:** 3.7/4.0

WORK EXPERIENCE

Flipside — Chicago, IL

Founder/CEO/ML Engineer — January 2017~

- Created, led AI Social Opinion Platform with 750+ users, 3-Person Technical Team, $33k raised
- Used Neural encoder, TextRank, and MPQA Dataset to train Main Opinion Extraction Algorithm and achieve 83% accuracy, a 60% improvement over standard NB-SVM
- Designed, Built Interactive Opinion Clustering Interface using Matrix Factorization, Dimension Reduction, Clustering, and Statistical Analysis
- Networking, Web Scraping, AWS, Database Management

Content Carnivores — Chicago, IL

Data Science Intern — June 2016-September 2016

- Designed, Built system that applied network theory algorithms to Twitter to speed up lead acquisition 25X
- Designed, Built Supervised Twitter Profile classifier for investors by engineering a lexicon, improved accuracy over existing system from 64% to 86%

Aavishkaar Co. Mumbai, India
Impact Investing Intern June 2015-August 2015

- Conducted Field Research with local small business owners in farms, vegetable markets, and local stores, culminating in analyst reports to understand ground reality in 3 industries
- Successfully pitched Hydroponic Farming startup to impact investment board

SELECTED RESEARCH/PROJECTS

Using Attention to Interpret Source Domains in Transfer Learning (In Progress)

- Independent Research Project interpreting deep domain transfer using self-attention
- Built Deep LSTM Language Model, LSTM Classifier, 2 Kinds of Attention, and weight sharing
- Advised by Karl Stratos at Toyota Technological Institute of Chicago

Big NLP on Reddit Comments

- Combined Sentiment Analysis Named Entity Recognition, and Knowledge Graph Search to extract (User, Public Figure, Sentiment) tuples and Unsupervisedly extract political groups
- Used MapReduce on Google Cloud Platform to analyze 26 GB dataset of Reddit Comments
- Extracted and averaged Word Embeddings to construct comment similarity matrix for clustering

Hybrid Recommender System for Movies based on Ratings and Scripts

- Jointly learned representations using Probabilistic Topic Modeling and Matrix Factorization to build SOA Recommender System
- Able to make high-quality recommendations on totally new content, solving "cold start" problem
- Crawled scripts and linked with IMDB Movie Reviews to create dataset of ratings and texts
- Improved over traditional collaborative filtering approaches by 6% percentage points

ACCOLADES

Social New Venture Challenge Winner
Polsky Accelerator Winner
Dean's List 2014-2017
Stanford BASES Finalist
Advanced 1st Place at Deerfield Debate tournament, largest tournament in New England Boarding School League
Music Director's Award at NMH School

SKILLS

Python Data Science Stack - PyTorch, sklearn, gensim, NLTK, Spacy, Networkx, Pandas, NumPy
(proficient) SQL, Javacript; (familiar) Scala, R, Java, C
Certified Advanced Proficiency in Mandarin, Conversational in Hindi
Plays Indian, African, Korean Latin Drums