

SIDDHARTH SACHDEVA

<https://sidsach.github.io/>

ssachdeva@uchicago.edu

6462494460

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

B.A. in Statistics, Minor in Computer Science

GPA: 3.7/4.0

Chicago, IL

Expected, June 2018

WORK EXPERIENCE

Flipside

Founder/CEO/ML Engineer

Chicago, IL

January 2017~

- Created, led AI Social Opinion Platform with 750+ users, 3-Person Dev Team, \$33k funding
- Designed, Built Main Opinion Extraction Algorithm supervised deep sentence embeddings, supervised subjectivity classification, and graph-based metrics of relevance
- Designed, Built Interactive Opinion Clustering Interface using Matrix Factorization, Dimension Reduction, Clustering, and Statistical Analysis
- Built Networking, Web Scraping, AWS, Database Management

Content Carnivores

Data Science Intern

Chicago, IL

June 2016-September 2016

- Designed, Built system that applied graph 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.

Impact Investing Intern

Mumbai, India

June 2015-August 2015

- Sourced, performed diligence for, and pitched Hydroponic Farming startup to investment board
- Conducted Field Research in farms, vegetable markets, and local stores, culminating in market research reports to understand ground reality in three industries

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, 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

\$20k Booth Social New Venture Challenge Winner

\$10k 2017 Polsky Summer Accelerator

Stanford BASES Venture Competition Finalist

1st Place at largest Debate Tournament in New England Prep School League

Music Director's Award at Northfield Mt. Hermon School

SKILLS

(Proficient) Python - PyTorch, sklearn, gensim, NLTK, Spacy, Networkx, Pandas, NumPy

(Familiar) Javascript, Scala, R, Java, C

Fluent in English, Mandarin, Conversational in Hindi

Plays Indian, African, Korean Latin Drums