

SIDDHARTH SACHDEVA
<https://github.com/siddsach>

RESEARCH STATEMENT

My research are in human-centered machine learning, specifically the ways in which natural language processing can be applied to human computer interaction and computational social science. This interests have motivated my research towards enabling natural language processing to overcome the barriers in the way of achieving this goal:

- Deep Transfer and Multi-Task Learning to improve data efficiency and generalization
- Improving model interpretability and mitigating bias.
- Interactive interfaces that improve these accessibility of these methods

EDUCATION

University of Chicago
B.A. in Statistics, Minor in Computer Science
GPA: 3.7/4.0

Chicago, IL
Expected, June 2018

WORK EXPERIENCE

Founder and CEO at Flipside January 2017~

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

Data Analytics Intern at Content Carnivores 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%

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

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

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

Fluent in English, Mandarin, Conversational in Hindi

Plays Indian, African, Korean Latin Drums