

WENCAO YANG

(812) 325 - 7652 ◇ yangwencao@gmail.com

linkedin.com/in/wencyang ◇ github.com/wencyang ◇ wencaoyang.medium.com

EDUCATION

Indiana University at Bloomington

Aug 2013 - Aug 2019

Ph.D. in Physics (GPA: 3.7/4)

Xiamen University

Sep 2009 - Jun 2013

B.S. in Physics (GPA: 3.7/4)

Coursework: Introduction to Statistics, Applied Linear Regression, Time Series Analysis

SKILLS

Languages Python, SQL, Java, R

Tools NumPy, SciPy, Pandas, Matplotlib, Scikit-learn, NLTK, SpaCy, Gensim, NetworkX, Keras, FastAPI, Flask, Docker, Airflow, Git, AWS (S3, RDS, EC2)

Knowledge Statistics, AB Test, Machine Learning, Neural Network, NLP, CV, Topological Data Analysis, Database, Object-Oriented Programming.

WORK EXPERIENCE

Strike Social

Oct 2019 - present

Data Scientist, Chicago, IL

- Built recommendation and optimization engines for advertisement campaigns, including building similarity (co-occurrence) matrices for different platform audiences, training regression models for campaign performance (cost-per-action) prediction, creating budget engines for budget distribution among different audience groups and across different platforms, and implementing optimization of pacing and bidding for advertisement campaigns.
- Categorized, filtered and ranked YouTube channels with NLP models (tf-idf, Word2vec, Doc2vec, BERT, Sentence-BERT, etc.) based on processed channel and video titles, descriptions, keywords (data enhancement + sentence embedding + clustering).
- Built data warehouse pipelines and analyzed key metrics from platforms of YouTube, Facebook etc. Built data visualization dashboards with SQL for internal and external users. Built Flask demos to demonstrate data insight results.

Indiana University

Aug 2013 - Aug 2019

Research Assistant, Bloomington, IN

- Analyzed polarized neutron spin echo scattering data and 2D/3D material characterization data. Fitted and visualized experimental data like nonlinear regression for certain spectrum and reciprocal space mapping for X-ray diffraction. Processed and analyzed low temperature transport measurement results.
- Published 10+ peer-reviewed publications with 160+ citations (Google Scholar).

SELECTED PROJECTS

Web Traffic Time Series Forecasting

Aug 2017 - Oct 2017

- A Kaggle competition to predict Wikipedia pages visiting numbers with historical data. Using a ‘median of medians model’ with a carefully picked list of sliding time windows to have an LSTM-like effect, achieving top 10% (Bronze medal) in the leader-board.

Network Community Detection

Oct 2016 - Jun 2017

- A research project to test state-of-art network (social graph clustering) community detection algorithm. Implemented Girvan–Newman algorithm with NetworkX and then tested stochastic block model with real social network data by inspecting the dependence of the degree of vertices inside a single community versus dependence of global degree of vertices.