Tianyu Zhang

https://tianyuz.me

Email : tianyuz2@andrew.cmu.edu / LinkedIn : johnzhangty
Mobile : +1 (650)505-8722 / +86 137-1859-3065

EDUCATION

- **Carnegie Mellon University School of Computer Science** — Pittsburgh, PA
  Bachelor of Science in Artificial Intelligence; GPA: 3.9/4.0, Dean's List. — *Expected May 2022*
  - **Relevant courses**: *Distributed Systems, *Algorithms Design & Analysis, *Scalable ML, Parallel Data Structures & Algorithms, AI: Representation and Problem Solving, Software Design & Concurrency, Theoretical CS

EXPERIENCE

- **ByteDance** — Beijing
  Software Engineering Intern — *June 2020 - Aug 2020*
  - **Live Streams Recommendation, Graph Embedding**:
    * Enhanced internal ML trainer (C++) performance by mitigating communication overhead. Reduced mini-batch forward latency by 40%+ in certain graph embedding training circumstances.
    * Extracted relations from Petabyte log data and built distributed user-author graphs using MapReduce. Devised graph encoders and trained an embedding of graph nodes using Tensorflow and ByteDance ML API.
    * Integrated end-to-end embedding to Click-Through-Rate prediction. Increased online user staytime (+3.5%), comment rate (+3.8%), and other metrics significantly in AB tests for Douyin (TikTok for Chinese market).
  - **Systems for Engineering Efficiency**:
    * Designed and developed a system from scratch in Django with RESTful APIs that creates and manages alerts for 100+ online models of 5 products and presents model health status on a dashboard. Attracted 70+ internal users and developers. Decreased the usual day-long response time to <1hr in a recent dataflow accident.
    * Constructed a pipeline that analyzes importance of 300+ features in a Monte Carlo fashion and performs feature modification on Terabyte model checkpoints on clusters based on results. Saved 2hr+ manual labor per iteration and 35k+ core-hour computing resources in total than hand-tuning.

- **Tsinghua Tongfang Nuctech** — Beijing
  Machine Learning Intern — *May 2019 - Jul 2019*
  - **Baggage Re-Identification for Smart Security Inspection**:
    * Updated existing model from verification scheme (Siamese nets) to the latest embedding scheme after implementing 5 models in PyTorch and evaluating performance on baggage datasets with hyperparam search.
    * Achieved 0.76 accuracy of CMC rank-1 on image retrieval from a 500-baggage gallery, improving baseline by 80.9% in accuracy and about 10x in speed.
  - **Framework Development**: Migrated research framework from C/Caffe to Python/PyTorch and developed additional features including activation visualization, checkpointing, and metrics platform integration.

- **Yuanfudao (EdTech Unicorn)** — Beijing
  Software Engineering Intern — *Jun 2017 - Jul 2017*
  - **Full-Stack Development**: Designed and developed from scratch a mobile event website in HTML, JavaScript, and Java that encourages high-school users to set up goals for new academic year and share to social networks. Reached 80k+ visits on event page and collected 10k+ responses.
  - **Data Analytics**: Performed regression analysis utilizing random forests on data of 8k+ students who improved Gaokao practice scores with the company's AI-enabled tutoring service. Summarized findings and generated proposals for team's marketing strategy over the next academic year.

PROJECTS

- **BitcoinMiner**: A failure-recoverable distributed Bitcoin miner with the homegrown Live Sequence Protocol that supports reliable communication under the server-client model. (Go)
- **TimeSeries**: An extensible analysis and visualization framework for time series data that allows plugins to provide custom data sources, analysis methods, and visualization methods. (Java(Swing))
- **Pop!**: A crowd-sourcing notification app that allows users to send signals into a group and all members of that group will be notified of the event in real-time (e.g. a pop quiz for 15-251). (JS(React)+Python(Django))
- **Zebrafish**: Pattern analysis and identification of stimuli-induced neural activity of a larval zebrafish using feature-based aggregation and distributed PCA. (Python(Spark))
- **LangDist**: Proposed a variant of the SIR model from epidemiology to predict future geographical distributions of language speakers. Generated proposals to inform global office site selection for multinational firms. (Python+MATLAB)
- **MagUnity**: An electromagnetism engine implementation for Unity (using Maxwell's equations) that enables game developers to integrate electromagnets and electron beams into gameplay. ($C^{\#}$/Unity, CASTIC 2017 Second Prize)

SKILLS

- **Systems**: Java, Go, C/C++, SQL, Spark/Hadoop, HDFS, RPC, Protobuf, Apache Kafka
- **Data & ML**: Python (pandas, sklearn, Tensorflow, PyTorch), MATLAB, Recommender Systems (Distributed LR, FFM, GraphSAGE), Reinforcement Learning (MDP, Q-learning), Computer Vision (CNN, ResNet, Attention)