HAITIAN (HILTON) JIANG

+86 136-8166-0596 jht2001120@outlook.com github.com/hiltonjiang

EDUCATION

Fudan University **Sep. 2019 – Jul. 2023 (expected)**

B.S. in Data Science *Shanghai, China*

- **GPA:** 3.85/4.00 **Ranking:** 2/80
- Teaching Assistant for Data Structure, Algorithm and Data Structure(Honor)
- Core Courses: Algorithm & Data Structure, Computer System, Numerical Methods, Optimization, Database, Stochastic Process, Computer Vision, Natural Language Processing, Parallel Computing

University of California, San Diego (UCSD) **Sep. 2021 – Dec. 2021**

Visiting Student, Computer Science and Engineering *San Diego, CA*

- **GPA:** 3.94/4.00
- Courses: Theory of Computation, Recommender System, Probabilistic Reason & Learning

RESEARCH EXPERIENCES

FreshGNN: Faithfully Refreshed Embeddings from a Selective History **Jun. 2022 – Present**

Deep Graph Library(DGL) group, Amazon Web Service(AWS) *Shanghai*

- Accelerated GNN training on large graphs by using historical cache to reduce computation and memory access.
- Designed check-in and check-out policies for selecting nodes staying in the cache and conducted experiments to compare the performance.
- Provided theoretical analysis for the convergence of training with history cache.
- Gained a 20x speedup with all optimizations on compared to the same GraphSAGE model using DGL on papers100M dataset, while achieving similar accuracy.

Scalable Graph Neural Networks using Subgraph Summarization **Jan. 2022 – Present**

Advisor: Prof. Zengfeng Huang *Fudan University, Shanghai*

- Used graph coarsening to summarize the subgraphs in order to speed up and regularize GNN training. The summarized subgraphs will have much fewer nodes while maintain the overall graph structure.
- Gained accuracy increase by about 1% in small datasets with common models like GCN
- Scaled up GNN training: implemented both pre-processing and training on multi-GPU; reduced the training time on papers100M from over 10 minutes to 27s for one epoch.

Recommendation System with Faithful Textual Explanation **Oct. 2021 – Present**

Advisor: Prof. Julian McAuley *University of California, San Diego*

- Evaluated current text-based recommenders through aspect-based sentiment analysis and found their text cannot faithfully explain their ratings because popular and positive aspects dominate the text.
- Designed a metric of faithfulness for recommendation explanation by learning and comparing the mapping from explanations to predicted ratings in order to measure the issue.
- Developed and trained a RoBERTa model to learn the underlying mapping between reviews and ratings on the Amazon Clothing and Google Local Restaurant dataset; got a MSE of 0.33 against 0.9 from typical models using user and item embeddings.
- Designed and implemented a new pipeline that first generate text explanations of recommendations and then use the text to predict faithful ratings; achieved similar MSE compared with traditional methods that cannot generate explanations.

PROJECT EXPERIENCE

Image Registration System | *Python, numpy* **May. 2022 – Jun. 2022**

- Built an image registration system using PyQT for GUI and numpy for implementing the stochastic gradient descent algorithm from scratch.
- Used optical-flow method for non-linear image transformation and different loss function for linear transformation: squared error, correlation coefficient, normalized cross correlation, mutual information and K-L divergence.
- Found the optical-flow method will break the global structure due to the coarse-to-fine procedure, so it should be used for local transformations; for the loss in linear transformation, K-L divergence is sensitive to parameters in this problem, while other loss functions all perform similarly good.

Few-shot Oracle Character Classification | *Self-supervised learning* **May. 2022 – Jun. 2022**

- Used sequence information extracted from images to generate augmented samples by adapting Sketch-BERT model.
- Added self-supervised learning on the direction prediction for the random rotation performed to the input image to enable better learning of representation; added ensembles on the feature vectors to make the network more robust.
- Gained about 15% higher accuracy than the SOTA method on all 1-shot, 3-shot and 5-shot tasks.

Network Service | *Go* **Sep. 2021 – Nov. 2021**

- Implemented a concurrent http server based on TCP sockets; can be visited through browsers or command line tools.
- Implemented a distributed file synchronizing client/server system based on consistent hashing for the load balance of server and remote procedure call for the client.

CPU Emulator with 5-stage Pipeline | *C* **Nov. 2020 – Dec. 2020**

- Implemented a RISC CPU with the classic five-stage pipeline, guard for hazard, pipeline with stall.
- Added support for data forwarding and branch instruction.
- Tested on a binary compiled from a MIPS assembly code containing loops and calculations and got the correct results.

PROFESSIONAL EXPERIENCE

Tencent Technology **Feb. 2021 – Jul. 2021**

IoT Security Intern *Shanghai, China*

- Collected, reproduced and exploited vulnerabilities(CVE) of IoT devices; conducted reverse-engineering on the firmwares.
- Implemented a crawler system using Scrapy and PostgreSQL to automatically gather information of the products and firmwares and store them in the database.
- Built an asynchronous downloader that fetches the source URL from the database and uses message queue(RMQ) for communication; used docker as the container.

HONORS & AWARDS

National Scholarship (TOP 1%) **Nov. 2021**

National Second Prize (TOP 3%), CUMCM **Dec. 2020**

(Contemporary Undergraduate Mathematical Contest in Modeling)

Second Prize Scholarship of Fudan University (TOP 10 %) **Nov. 2020**

SKILLS

Languages: Mandarin (Native speaker), English (Fluent)
Programming: Python, C/C++, Go, CUDA, MATLAB, R, SQL
Frameworks: Linux, Git, PyTorch, GPU programming, Spark, Docker, Web Crawler(Scrapy)