

ChenyuanWu

E-mail: wucy@seas.upenn.edu | Website: chenyuanwu.com

EDUCATION

University of Pennsylvania

2020 - Present

Ph.D. in Computer Science, School of Engineering and Applied Science

- Advisor: Prof. Boon Thau Loo

Beijing Jiaotong University

2016 - 2020

B.E. in Communication Engineering, School of Electronic and Information Engineering

- General GPA: 3.9/4.0
- Thesis: Efficient Retrieval and Data Mining of Trajectory Big Data
- Advisor: Prof. Yishuai Chen

PAPERS

[1] Shaobo Wang, Hui Lyu, Jiachi Zhang, **Chenyuan Wu**, Xinyi Chen, Wenchao Zhou, Boon Thau Loo, Susan B. Davidson and Chen Chen, *Provenance for Probabilistic Logic Programs*, International Conference on Extending Database Technology (EDBT), 2020. **Best Paper Award**.

[2] Haoxian Chen, **Chenyuan Wu**, Manqiu Zhang, Yahui Li, Zhijia Chen, Boon Thau Loo, Mayur Naik and Mukund Raghothaman, *Synthesizing Verifiable DSL Programs from SDN Controller Applications*, ACM Symposium on Cloud Computing (SoCC), 2020. Status: Under Review.

RESEARCH INTERESTS

I am interested in practically applying machine learning to networked systems, seeking to secure better robustness and performance for systems, and higher efficiency for operators. My research covers the broad areas of networked systems, including protocols, fault detection, and software defined networking.

AWARDS & HONORS

President's Scholarship in Beijing Jiaotong University (10 out of 15000)	Nov. 2019
Chinese National Scholarship	Sep. 2019
Chinese National Scholarship	Sep. 2017
Second Prize in the National College Physics Competition	Dec. 2017
First Prize in the Mathematics Competition of Beijing Jiaotong University	Jun. 2017

WORKING EXPERIENCES

Microsoft Research

Contractor, Mobility and Networking Group

Apr. 2020 – Jun. 2020

- Crawled Pantheon's archived data, and patched Pantheon's wrappers and third-party libraries to provide suitable API for web downloads and video streaming.

Huawei, Central Research Institute

Research Intern, Network Technology Lab

Nov. 2019 – Jan. 2020

- Designed and validated a mechanism for bottleneck routers to stabilize the network performance, when facing a diverse mix of long and short flows.
- Measured and analyzed the performance of multipath transport protocols in heterogenous networks.

RESEARCH EXPERIENCES

Synthesizing Verifiable DSL Programs from SDN Controller Applications

July. 2019 – Sept. 2019

Research Assistant, Distributed Systems Lab, University of Pennsylvania;

Advisor: Prof. Boon Thau Loo

Objective: AutoDSL is a toolchain that synthesizes SDN controller implementations in domain specific languages (DSL) from input-output execution traces, aiming to bridge the gap between legacy and DSL implementations. AutoDSL uses code analysis to extract internal states from legacy SDN implementations, and syntax-guided synthesis to generate logic programs that are consistent with the input-output example traces. The logic programs are compiled into simpler, more concise logic-based DSL implementations, which can then be used as a basis for correctness verification.

- Implemented two compiler modules of the toolchain in Scala, which could automatically translate the intermediate logic programs into Flowlog rules and dynamic policies in Pyretic;
- Implemented 7 legacy programs on the FloodLight platform, and implemented an execution traces collector for FloodLight;
- Compiled logic programs of MAC learning and L2/L3 firewalls into DSL implementations, validated their functionality in Mininet;
- Compared the end-to-end difference in throughput between handwritten and synthesized DSL implementations by running perf tests over a FatTree topology in Mininet, compared their difference in response ability by running Cbench tests;
- Submitted a paper to ACM SoCC [2].

Provenance for Probabilistic Logic Programs

July. 2019 – Sept. 2019

Research Assistant, Distributed Systems Lab, University of Pennsylvania;

Advisor: Prof. Boon Thau Loo

Objective: P3 is a novel provenance model and system for analyzing probabilistic logic programs. P3 enables four kinds of queries: traditional explanation queries, queries for finding the set of most important derivation within an approximate error, most influential queries, and modification queries that change the output probability with fewest modifications to input data or program.

- Implemented two parallel computing methods for the most influential query in P3, where each work unit is in charge of a single round of Monte Carlo simulation or computing the influence value of a single Literal;
- Set up parallel computing environments on both Nvidia GPU and Intel CPU, run experiments on large provenance graphs, and analyzed results from different methods and different hardware settings, boosting the performance by more than 20-fold faster than the sequential method at most;
- Conducted a Visual Question Answering case study, further enabling the system to find inappropriate input data and show the modification strategy with minimum cost in order to make the prediction right;
- Published a paper in EDBT [1].

Prostate Image Segmentation based on Machine Learning

Sept. 2018 – Feb. 2019

Independent Project, Institute of Intelligent Sensing and Information Processing (National Teaching Base for Electrical and Electronic Engineering), Beijing Jiaotong University;

Advisor: Prof. Yahui Peng

Objective: Separate different tissue structures, especially the prostate tissue, by applying Gaussian Mixture Model and the Nearest Neighbor method to multimodal images.

- Data Preprocessing: For every given slice, programmed to find the slice with the shortest distance and the same normal direction in each set of slices with different modes;
- Feature Engineering: Did sparse sampling with equal spacing on 2D plane, and the different gray values of each sample point were used as features to form a multi-dimensional vector space;
- Used GMM to cluster the sample points in the vector space, while the remaining unclassified pixels in the 2D plane were mapped to the vector space and labeled by leveraging kNN;
- Repeated the above process for a case of multimodal 3D volume, used the 3D slicer software for stereo reconstruction, visualizing the 3D tissue structure after segmentation, and compared the results with those under 2D condition.

TEACHING & ACTIVITIES

- **Conference Reviewer:** served as the secondary reviewer of the International Conference on emerging Networking Experiments and Technologies (CoNEXT 2020).