Xiaoyuan Liu (刘啸远)

Undergraduate in Computer Science

Email:	lxy9843@sjtu.edu.cn
Tel:	(+86) 132 6229 0218
Home Page:	https://littleround.cn/about/

EDUCATION

Shanghai Jiao Tong University **Shanghai, China**
Honors Bachelor of Science (B.Sc. Hons) in Computer Science Sept. 2016 ~ June 2020 (Expected)

- Member of **ACM Honors Class**, which is an elite CS program for top 5% talented students
- Average score: 90/100

HONORS AND AWARDS

The 32nd China's National Olympiad in Informatics (NOI) Silver Medal	2015
KoGuan Encouragement Scholarship (**Top 0.3%**, SJTU)	2017
Zhiyuan Honorary Scholarship	2016, 2017, 2018
Academic Excellence Scholarship	2017, 2018, 2019
Outstanding Student Cadre (**Top 0.8%**, SJTU)	2018

RESEARCH EXPERIENCE

Visiting Student Researcher **Berkeley, CA, USA**
University of California, Berkeley, advised by Prof. Dawn Song July 2019 ~ Dec. 2019 (Expected)

- LASER: **L**earning to **A**utomate **S**ocial **E**ngineering **R**esistance
  - LASER project focuses on the design of an automated attack detection and attacker identification system, able to conduct active investigation of social engineers using dialogue system technology.
  - Key contributor to the Berkeley LASER project. Developed a phishing email detector based on state-of-the-art NLP techniques like the BERT model. It achieved 100% recall and 0.7% false positive rate on industrial dataset utilizing meta-data, body part and attachment features.
  - Designed an attack response system based on semantic similarity to acquire PII of attackers.
  - Designed a scalable, fault-tolerant distributed framework to tie together all modularized and containerized system components, which stayed operational and processed thousands of emails during a half month online evaluation using real-world data.
  - Managed the code maintenance and the deployment on k8s cluster infrastructure. Provided a list of RESTful APIs and a python SDK to support easy third-party development. Wrote detailed documentation to support the code integration for collaborators from other universities.
- Automated threat hunting over system audit logging using cyber threat intelligence
  - The project focuses on the design and development of a novel system to automate the audit logging-based forensic investigation of sophisticated cyber-attacks using threat intelligence. It contains an attack behavior extraction component that extracts knowledge from natural-language security articles to generate graph representations and a query engine component that uses domain-specific language to conduct threat hunting in an optimized database of system audit logging events.
  - Developed the attack behavior extraction component using named entity recognition, relation extraction, coreference resolution, and other related knowledge graph construction techniques.
- Text-to-SQL generation
  - Improved the performance of natural language to SQL generation by leveraging meta learning training method.
  - Purposed a new way to evaluate the adaptability of Text-to-SQL models for unseen database schemas.
- Measurement of language model robustness
  - Build a consistent framework to run experiment using language models like word2vec, glove, BERT, RoBERTa and related network encoders like CNN, LSTM, Transformers.
  - Measured the robustness of trained language models by testing it on a relevant domain with distributional shift on different tasks like sentimental analysis, sentence similarity, QA, etc.

Undergraduate Researcher **Shanghai, China**
Shanghai Jiao Tong University, advised by Prof. Kai Yu July 2018 ~ June 2020 (Expected)

- Reinforcement learning for task-oriented dialogue management
  - In this work, our group proposed a novel structured actor-critic approach to implement structured deep reinforcement learning (DRL), which not only can learn parallelly from data of different dialogue tasks but also achieves stable and sample-efficient learning.
  - Developed a multi-domain dialogue environment by combining existing single-domain user simulators while maintain the consistency of the dialogue.

- Speech tone classification
  - Built a classifier for tones of single Chinese characters. By analyzing the f0/energy sequences using a set of well-designed rules, achieved an accuracy above 99% in a multi-class classification setting.
  - Won first place in kaggle competition hold by AISPEECH.

SELECTED PROJECTS

RL Framework for Image Classification Fooling ■ Python 2018
Reinforcement Learning, Model Robustness

- Coursework of *"Frontiers of Computer Science"*
- Proved that it is possible to fool image classifiers **in the black box setting** using RL techniques.

Reinforcement Learning in the Card Game Dou Di Zhu ■ Python 2019
Hierarchical Reinforcement Learning, Backend

- Coursework of CS492 *"Reinforcement Learning"*, won **first** place in class.
- Investigated the Chinese card game Dou Di Zhu, an imperfect information game with randomness.
- Implemented several rule-based baseline agents which have human-compatible performance.
- Showed that a hierarchical reinforcement learning agent using summary actions can benefit from the ability of making high-level decisions and outperform all baselines.

Mx* Compiler ■ Java 2018
Assembly Language, Code Generation and Optimization, ANTLR

- Coursework of *"Compilers"*
- Developed a compiler that compiles C-and-Java-like language (Mx*) to NASM.
- Implemented optimizations like constant replacement, function inline and loop unrolling.

QuPlayground ■ JavaScript 2018
Quantum Computing, Simulation, UI Frontend

- Coursework of *"Quantum Information Science"*
- Built a quantum computation simulator from scratch with almost no dependency.
- Designed a convenient and intuitive GUI using GoJS to help user construct and demonstrate their quantum circuits. Examples include Bell test, quantum teleportation and Shor algorithm.

Toy ML System ■ C++, Python 2017
Machine Learning System, CUDA Programming, Dynamic Library

- Designed a TensorFlow-like machine-learning system which support simple operators including matmul, dropout, softmax & relu, conv2d & max_pool with autograd.
- Supported optimizers like vanilla gradient descend and ADAM. Utilizes a carefully written multi-thread C++ dynamic library to accelerate the computation of convolution and max pooling operation.

RISC-V CPU ■ Verilog 2018
Computer Architecture, Tomasulo, FPGA Programming

- Designed a RISC-V CPU that supports RV32I Base Integer Instruction Set V2.0 (2.1~2.7).
- Designed a modified Tomasulo structure to support superscalar with arbitrary number of ALUs.

Chinese Land Battle Chess AI ■ C++ 2016
Game Theory (Minimax), Alpha-beta Pruning, Genetic Algorithm

- Built a rule-based AI for Chinese Land Battle Chess. Adapted techniques like alpha-beta pruning, beam search and time estimation to guarantee the searching time for each step within 1 second limit.
- Designed a genetic algorithm to screen for a better initial arrangement of the chess pieces.
- Won **second** place in the round-robin tournament in class.

TEACHING EXPERIENCE

Lead Teaching Assistant *C++ Programming (A)* Fall 2017
Lead Teaching Assistant *Data Structures* Spring 2018
Student Instructor *Principle and Practice of Computer Algorithms* Summer 2018

ACTIVITIES

Student Council Vice President 2018
Head of the Department of Culture and Sports, Student Union 2017

SKILLS AND INTERESTS

Languages: Mandarin (Native), Japanese (Beginner)
Programming: C++ / Python / Java / JavaScript / Verilog / MATLAB / Pascal
Technical experience:

- Web: Django / Flask / Express / Koa / Jade (Pug) / Swagger
- System & Database: Mininet / Docker / Kubernetes / Jenkins / MySQL / MongoDB / Redis
- Other: LaTeX / Markdown / Wireshark / Qt / Wayland & Weston / Vivado

Interests: Photography, Badminton, Image & Video Editing