Linhai Ma

+1-919-884-0535 | rinnkai.ba2019@gmail.com | https://www.linkedin.com/in/linhai-ma-6441861bb/ | https://sarielma.github.io/

Introduction

He is a final-year Ph.D. student of Computer Science at the University of Miami. He has over three years of research and application experience in machine learning, deep learning, and medical signal/image analysis. He also has industry experience as an intern machine learning engineer in recent years. In addition, he has a solid background in computer science and programming, and research experience in concurrent C++ software testing.

Education

University of Miami, Coral Gables, FL 2017/08-2023/12
Ph.D. in Computer Science, GPA: 3.85/4.0

Institute of Software, Chinese Academy of Sciences, Beijing, China 2014/09-2017/07
Master of Science in Computer Science, GPA: Top 10%

Northeastern University, Shenyang, China 2010/09-2014/06
Bachelor of Science in Information Security, GPA: Top 5%
Courses: C\C++, Java, Data Structure, Computer Architecture, Operating System, Computer Network, etc.

Skills

Programming Languages:	Python, C/C++, Java, C#, Ruby, R, SQL, Prolog, etc.
Tools:	Pytorch, Pandas, Scikit-learn, Numpy, PyQT5, Java Swing, etc.

Industry Experience

Cadence Design Systems, Inc., Simulation R&D Team, San Jose, CA 2022/05-2022/08
Intern Software Engineer in Machine Learning

- Designed a machine learning system to accelerate the evaluation of worst-case measurement of Integrated Circuit (IC), which normally requires thousands of IC simulations, using Python, Pytorch, etc.
- Sampled simulation inputs from the input space using Latin Hypercube Sampling and performed simulations on these inputs to get simulation outputs, which form the initial training set to train a Gaussian Process regression model.
- Selected the next simulation input that most likely leads to the worst-case measurement using Bayesian optimization. Put this input with its simulation output into the training dataset for the next epoch of training.
- Performed forward feature selection to eliminate noisy input features that lead to outlier output of the simulation.
- Reduced the time needed to discover the worst-case measurement by over 90%.

Cadence Design Systems, Inc., Simulation R&D Team, San Jose, CA 2021/05-2021/08
Intern Software Engineer in Machine Learning

- Designed a machine learning system to predict the output of Integrated Circuit (IC) simulation via Python, Sk-learn, etc.
- Wrote a data pipeline parsing and preprocessing data from previous simulations of this evaluation to get the training set.
- Selected Random Forest to predict, which accommodated the highly diverse ranges of simulation input features.
- Developed a UI to visualize the impact of each feature of the simulation input on the output via PyQT5 and Matplotlib.
- Reached a prediction accuracy of a mean absolute error of 0.99 picoseconds.

Hillstone Networks, Inc., Beijing, China 2015/06-2015/08
Intern C\C++ Developer

- Wrote URI filters in the firewall machine to filter out abnormal requests using C\C++, GDB, etc.
- Developed an automatic SQL manipulator to manipulate the SQLite database according to client requests using Ruby.

Huaxin Education Technology Co., Ltd., Shenyang, China 2013/06-2013/08
Intern Java Developer

- Developed a vulnerability scanner to discover the vulnerability of the web pages using Java, JSP, Oracle, etc.
- Developed the UI for a better user experience using Java Swing.

Research Experience (Please find more details of the research projects on my page https://sarielma.github.io/)

University of Miami, Coral Gables, FL 2019/09-now

Research Assistant

Topic: Adversarial robustness of deep neural networks and medical signal/image analysis

● Designed a CNN model for a variant-length, 12-lead, 9-class electrocardiogram (ECG) classification task from the China Physiological Signal Challenge 2018 using Python, Pytorch, etc., which achieved top-6 performance in the challenge.

● Proposed a regularization method aiming to penalize the estimated upper bound of noise-to-signal ratio in the output of the deep neural networks (DNNs) in order to improve the resistance of DNNs against white noises and adversarial noises.

● Designed an adversarial training algorithm by adapting the margin of each training sample to improve DNNs for better adversarial robustness.

● Developed an algorithm based on loss-defined margin to improve the robustness of DNNs on various medical image applications, including MRI image segmentation, Cephalometric landmark detection and blood cell object detection.

University of Miami, FL and Northwestern University, IL 2018/01-2019/09

Research Assistant

Topic: Social network analysis and data mining

- Discovered which evolution pattern is more likely to lead to a successful academic group via a data mining approach.
- Parsed author and publication data from 1991 to 2018 from Microsoft Graph database via SQL, and built up million-level coauthor networks for each year, using Python, NetworkX, etc.
- Designed a Monte Carlo-based clustering algorithm to cluster the graph into author groups, via Python and R.
- Defined these author groups' evolution patterns (splitting, merging, etc.) in two adjacent years.
- Made statistics on frequency of each group's evolution patterns and the group's successfulness (e.g., the number of citations) to conclude which pattern contributes more to the successfulness of each group, via Python and R.

Institute of Software Chinese Academy of Sciences, Beijing, China 2015/07-2017/06

Research Assistant

Topic: C++ concurrent software testing

- Developed a generator to generate multi-thread C++ test cases based on the concurrent C++ data structure to be tested.
- Proposed three adaptive strategies to improve the C++ test case generation, which discovered up to 6% more potential concurrent errors and reduced the time cost by up to 10%.

Project

Project: Sentiment analysis on review texts from Amazon

- Used the customers' review texts to predict their ranks (Star 1 to Star 5) on the corresponding product.
- Encoded the review texts with Vector Space Model (VSM) with Tf-idf and Latent Dirichlet Allocation (LDA) into feature vectors via Python, Java, MALLET, WEKA, and Windows Batch Script.
- Used the encoded review texts and the corresponding customers' ranks (Star 1 to Star 5) to train classifiers, e.g., SVM and Naïve Bayes.
- Predicted a new customer's rank on a product, given his/her review text.
- Reached a prediction accuracy of 70% on testing set.

Academic Reviewer Experience

- International Conference on Machine Learning 2022 (ICML2022) – AI top conference
- Neural Information Processing Systems 2022 (NeurIPS 2022) – AI top conference
- Scientific Programming – A peer-reviewed journal in software engineering
- IEEE Journal of Biomedical and Health Informatics (JBHI) – A peer-reviewed journal in biomedical informatics
- Expert Systems with Applications (ESWA) – A peer-reviewed journal in intelligent systems applications
- Artificial Intelligence (AI) – A peer-reviewed journal in AI
- Computers in Biology and Medicine (CIBM) – A peer-reviewed journal in biomedical informatics
- Computer Systems Science and Engineering – A peer-reviewed journal in computer systems science
- Computers, Materials & Continua (CMC) – A peer-reviewed journal in computational materials science and engineering

Publications

- Linhai Ma, Liang Liang. "Increasing-Margin Adversarial (IMA) Training to Improve Adversarial Robustness of Neural Networks." https://arxiv.org/abs/2005.09147. Computer Methods and Programs in Biomedicine (2023).

- Linhai Ma, Liang Liang, " Improving Adversarial Robustness of Deep Neural Networks via Adaptive Margin Evolution." Neurocomputing (2023).

- Linhai Ma, Liang Liang, "Towards lifting the trade-off between accuracy and adversarial robustness of deep neural networks for medical image classification and segmentation." Medical Imaging 2023: Image Processing. Vol. 12464. SPIE, 2023.

- Linhai Ma, Liang Liang. "A Regularization Method to Improve Adversarial Robustness of Neural Networks for ECG Signal Classification." Computers in Biology and Medicine (2022).

- Linhai Ma, Liang Liang. " Enhance CNN Robustness Against Noises for Classification of 12-Lead ECG with Variable Length." 19th IEEE international conference on machine learning and applications (ICMLA 2020).

- Linhai Ma, Liang Liang. "Improve robustness of DNN for ECG signal classification: a noise-to-signal ratio perspective." International Conference on Learning Representations (ICLR 2020) Workshop AI for Affordable Health.

- Linhai Ma, Peng Wu, Tsong Yueh Chen. "Diversity driven adaptive test generation for concurrent data structures." Information and Software Technology (2018).

- Linhai Ma, Liang Liang, "Adaptive Adversarial Training to Improve Adversarial Robustness of DNNs for Medical Image Segmentation and Detection." https://arxiv.org/abs/2206.01736. To be submitted.

- Liang Liang, Linhai Ma, Linchen Qian, Jiasong Chen. " An Algorithm for Out-Of-Distribution Attack to Neural Network Encoder." https://arxiv.org/abs/2009.08016. To be submitted.