Linhai Ma

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

Education

University of Miami, Coral Gables, FL 2017/08-2023/12
Ph.D. in Computer Science, specializing in Deep Learning and Medical Image Analysis, GPA: 3.85/4.0

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

Northeastern University, Shenyang, China 2010/09-2014/06
Bachelor of Science in Information Security, GPA: Top 5%

Skills

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

Industry Experience

Cadence Design Systems, Inc., San Jose, CA 2022/05-2022/08
Intern Software Engineer

- Designed a machine learning system to predict the worst-case measurement of Integrated Circuit (IC) in one IC simulation process. (Simulation is a process to verify the performance of IC. One simulation has one input feature vector and one output measurement vector. Typically, thousands of simulations are needed in one IC simulation process to discover the worst-case measurement.)
- Built a Gaussian Process model to predict, with previous simulation results in this IC simulation process.
- Used Latin Hypercube Sampling to sample simulation inputs from the input space. Performed simulations on these inputs to get simulation outputs. Used these simulation inputs and outputs to form the initial training set to train the model.
- Used Bayesian optimization to select one next simulation input from the input space that most-likely leads to the worst-case measurement. Put this input with its simulation output into the training dataset for the next epoch of training.
- Performed forward selection to eliminate noisy input features that lead to outlier output of simulation.
- Reduced the time needed by an IC simulation process to discover the worst-case measurement by over 90%.
- Skills used: Python, Pytorch, Botorch, Scikit-learn, Pandas, Linux (red hat), etc.

Cadence Design Systems, Inc., San Jose, CA 2021/05-2021/08
Intern Software Engineer

- Designed a machine learning system to predict the output of new Integrated Circuit (IC) simulation with results from finished simulations in one IC simulation process.
- Used the tree-based regression model, Random Forest, and Gradient Boost Decision Tree to predict, because the ranges of each feature of the simulation input are highly diverse.
- Implemented a data pipeline actively parsing the results from finished simulations in this IC simulation process and putting the preprocessed data into training set. Used one-hot encoding to encode the categorical features in the simulation input.
- Developed a user interface to visualize the relationship between each feature of the input of simulation and the output of simulation.
- Reached a prediction accuracy of a mean absolute error of 0.99 picoseconds, which is very satisfying.
- Skills used: Python, Xgboost, Scikit-learn, Pandas, Numpy, Scipy, Matplotlib, PyQT5, Linux (red hat), etc.

Hillstone Networks, Inc., Beijing, China 06/2015-11/2015
Intern Software Engineer

- Modified and Migrated an Intrusion Prevention System (IPS) server to a new server machine.
- Developed a packet sniffer periodically capture packets from the server.
- Developed an automatic SQLite manipulator to automatically manipulate the SQLite database according to client requests.
- Developed HTTP URI filter in the firewall machine to filter out abnormal requests.
- Skills used: C\C++, Python, SQL, Ruby, Ruby on Rails, Linux (Centos) etc.

Huaxin Education Technology Co., Ltd., Shenyang, China 06/2013-08/2013
Intern Software Engineer

- Developed a Vulnerability Scanner for web applications to discover the vulnerability of the web pages against attacks, such as SQL Injection, Cross-Site Scripting vulnerabilities.
- Developed the user interface for a better user experience.
- Skills used: Java, JSP, Oracle, Java Swing, etc.

Research Experience

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

Research Assistant

Project: Improve the robustness of deep-learning-based automatic electrocardiogram diagnosis

- Designed a CNN model for a variant-length 12-lead electrocardiogram classification task (from the China Physiological Signal Challenge 2018) which achieved top-6 performance in the challenge.
- Designed a multiple-layer perceptron (MLP) model for electrocardiogram data classification (from PhysionNet's MIT-BIH dataset), whose prediction accuracy is 92%.
- Proposed a regularization method to improve the deep-learning-based electrocardiogram automatic diagnosis for better robustness.
- Skills used: Python, Pytorch, Pandas, Scikit-learn, etc.

Project: Increasing-margin adversarial training to improve adversarial robustness of neural networks

- Proposed a training method via increasing margin ideas to improve the neural networks for better robustness and gave the theoretical proof of the training method's convergence.
- Evaluated the proposed method with residual net classifiers on public image datasets, e.g., CIFAR10, TinyImageNet, etc., and real-world medical images, e.g., Covid-19 CT.
- The proposed method significantly improved the model's robustness against adversarial noises with minimal accuracy degradation.
- Skills used: Python, Pytorch, Pandas, Scikit-learn, etc.

Project: Improve variant deep-learning-based medical applications for better robustness

- Proposed adversarial training methods to improve the following deep learning applications for better robustness:
    - Unet-based model (nnUnet) for Heart, Hippocampus and Prostate MRI image segmentation.
    - Multi-task Unet-based model for cephalometric landmark detection.
    - YOLO V5 for blood cell detection.
    - Transformer-based model (TransUnet) for abdominal organ segmentation.
- Improved the accuracy of these models under testing noises.
- Skills used: Python, Pytorch, Pandas, Scikit-learn, etc.

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

Research Assistant

Project: Discovering what contributes to a successful academic group via a data mining approach

- Parsed author and publication data (from 1991 to 2018) from Microsoft Graph database to build up one author-to-author graph (millions by millions matrix) for each year.
- Designed a Monte Carlo-based clustering algorithm to cluster the graph into author groups.
- Defined these author groups' evolution patterns (splitting, merging, etc.) in two adjacent years. Collected these patterns to show how each author group evolves over the years.
- Take statistics on times of occurrences of each group's evolution patterns and the group's successfulness (e.g., the number of citations) to conclude what evolution pattern is more likely to contribute to a successful group.
- Skills used: Python, R, NetworkX, SQL, etc.

State Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences, Beijing, China 07/2015-06/2017

Research Assistant

- Designed and developed a C++ generator to automatically parse the C++ class (the concurrent data structure implemented with the pthread library) and to generate multi-thread test cases that invoke this C++ class.
- Proposed three adaptive algorithms 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%.
- Skills used: C\C++, Python, etc.

Course Project

Project: Sentiment analysis on review texts from Amazon

- Analyzed the sentiment inside the review texts and predicted how buyers felt about the things they bought from their review texts.
- 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 SVM and Naïve Bayes via Scikit-learn to classify the review texts into five categories, corresponding Star 1 to Star 5, to show their potential sentiment inside the review texts.
- Showed that the performance of LDA is as good as VSM in this task.

Reviewer Experience

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

Publications

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

- Linhai Ma, Liang Liang, " Improving Adversarial Robustness of Deep Neural Networks via Adaptive Margin Evolution." Under review.

- 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. Under review.

- Linhai Ma, Liang Liang. "Increasing-Margin Adversarial (IMA) Training to Improve Adversarial Robustness of Neural Networks." https://arxiv.org/abs/2005.09147. Under review.

- 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. Under review.

- Linhai Ma, Liang Liang. "A Regularization Method to Improve Adversarial Robustness of Neural Networks for ECG Signal Classification." Computers in Biology and Medicine 144 (2022): 105345.

- 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 103 (2018): 162-173.