Rangeet Pan

rangeet@iastate.edu www.linkedin.com/in/rangeetpan https://rangeetpanisu.web.app/

Research Statement

With the current increases in Machine Learning (ML) usage, there is a potential technical debt that has emerged in terms of privacy, fairness, and robustness. The obvious question arises, how can we repay those debts? My research is to incorporate program analysis techniques used in software engineering into ML and overcome various problems by exploring ML models statically. My focus area of research is to study deep neural networks (DNN) to understand the characteristics of the system and increase robustness by exposing the DNN bugs, creating defense against adversarial attacks, etc.

Education

Iowa State University	Ames, USA	Computer Science	Ph.D.	2018-Current
University of Houston	Houston, USA	Computer & Systems Engineering	M.S.	2016-2018
Heritage Institute of Technology	Kolkata, India	Computer Science & Engineering	B.Tech	2008-2012

Experience

Aug 2018 – Current
Research Assistant, Iowa State University

- Work in the Laboratory of Software Design under the supervision of Dr. Hridesh Rajan.
- Working in static program analysis to study the characteristics of the deep neural network model for achieving robustness.

Aug 2016 – May 2018
Research Assistant, University of Houston

- Prepared end-of-course faculty evaluation reports for each academic department, including extracting data with SQL and evaluating data with SPSS
- Statistical calculation using Python and visualize the same using tableau.
- Conducted customized statistical analysis, for example, Point- Biserial Correlation Analysis for exam questions, per instructor's request to help instructor determine the quality of exam.

March 2013 – July 2016
Test Engineer, Infosys Limited

- Well-versed in all stages of Software Development and Testing Lifecycle (SDLC/STLC).
- Queried and tested RDBMS such as Oracle, MS SQL Server using SQL, PL/SQL for data integrity.
- Proficient in using Python packages – NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, Spark MLlib for data wrangling and analysis.
- Performed Data profiling and mapping, created class diagrams and ER diagrams and used SQL queries to filter data within database.
- Proficient in Project Management Methodologies like Agile (SCRUM) and Waterfall.
- **Accomplishment**: Client Aetna achieved CMMI Level 3 for the year 2015. Reduced manual effort and resulted in savings of $100k-150k/annum. Awarded FAME Awards and recognized across the company as the second-best in-house tool. Awarded INSTA for sharing an idea in Zero Distance initiative.

Publications

- Md Johirul Islam, Giang Nguyen, Rangeet Pan, and Hridesh Rajan, "A Comprehensive Study on Deep Learning Bug Characteristics," ESEC/FSE'19: The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, August 2019.
- Pan, Rangeet. "Static deep neural network analysis for robustness." In Proceedings of the 2019 27th ACM Joint

Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 1238-1240. ACM, 2019.

- Islam, Md Johirul, Hoan Anh Nguyen, Rangeet Pan, and Hridesh Rajan. "What Do Developers Ask About ML Libraries? A Large-scale Study Using Stack Overflow." arXiv preprint arXiv:1906.11940 (2019). (Under Review TSE'19)
- Islam, Md Johirul, Rangeet Pan, and Hridesh Rajan. "Repairing Deep Neural Networks: Fix Patterns and Challenges.", *ICSE'20: The 42nd International Conference on Software Engineering*, May 2020.
- Pan, Rangeet, Md Johirul Islam, Shibbir Ahmed, and Hridesh Rajan. "Identifying Classes Susceptible to Adversarial Attacks." arXiv preprint arXiv:1905.13284 (2019).

Projects

- **Type Independent Features on Lambda Calculus**
  In this study, we chose five different such features, Enumerator, Packing, Unpacking, Zip and Unzip. These features are all iterable features and can be implemented on top of lambda cal- culus, list, pair and basic language operations.
- **Hurricane path prediction from HURDAT dataset**
  Every year hurricanes cause damages in the USA and all around the world at a huge scale. Hence, any prior knowledge of a hurricanes path will facilitate in predicting the region where it might be headed next and help in taking precautionary measures beforehand. We can exploit the dataset named HURDAT database, provided by the National Oceanic and Atmospheric Administration (NOAA) to feed it into different supervised learning models. Subsequently, a comparison of the accuracy of the result predicted is discussed.
- **Flight Data Analysis Using Spark**
  Analysis of 15 years US domestic flight data using spark in different input format (csv, parquet, sequence file, json). Different statistical calculations are done upon the dataset using PySpark and then comparison of execution time of each iteration of different input format.
- **Vehicle Detection and Tracking**
  Detecting car in a video sequence and tracking the car in the video sequence. A classifier is trained for detecting car in an image and the neural network using for tracking the car in the sequence using sliding window technique.

Achievements

- Won Merit Scholarship, Phi Beta Delta organization, University of Houston, 2018.
- Coursera certificate on Neural Network and Deep Learning, Machine Learning, 2018.
- Developed the judging website for the "Science and Engineering Fair of Houston" and "Mar Rover", University of Houston, 2017.
- Awarded Dean's Scholarship, University of Houston, 2017.

Applications & Programming

- **Applications:** MATLAB, IBM SPSS, Hadoop, Spark, Dreamweaver, Visual Studio
- **Programming**: Python, MATLAB, Java.

Courses

- **Data Analytics:** Big Data Analysis, Artificial Intelligence.
- **Programming and Design:** Advanced Topics in Programming Languages, Programming Language Design and Semantics, Software design, Design and Analysis of Algorithms, Theory of Computation, Introduction to cybersecurity.