Yiqian Liu

- • - • Vancouver, BC
https://www.linkedin.com/in/yiqian-liu/

Curriculum Vitae

RESEARCH INTERESTS

Obtained degrees in Computer Science and Mathematics, I am a working Data Scientist with published paper, filed patent, and various experience of machine learning and R&D projects. **Computer vision** is my most recent focus, as demonstrated by the personal project on object recognition with YOLO. I also found computer graphics interesting. Autonomous vehicles (ambulances) and diagnoses aided by computer graphics or image analyses are the two applications that I care the most.

EDUCATION

Master of Computer Science, thesis option — 2013 – 2016
University of New Brunswick — Fredericton, NB
Thesis: Machine Learning for Wind Power Prediction. Supervisor: Huajie Zhang
Relevant Courses: Machine Learning & Data Mining, Foundations of AI, Computer Graphics

Bachelor of Science in Mathematics — 2009 – 2013
Shandong University — Jinan, China
Visiting Student at Loughborough University, England, 2012 – 2013

PUBLICATIONS

Yiqian Liu, Huajie Zhang, "An Empirical Study on Machine Learning Models for Wind Power Predictions", Proceedings, 15th IEEE International Conference on Machine Learning and Applications (IEEE ICMLA), Dec. 2016.

- Attended conference and presented poster in Anaheim, California, USA, December 18-20, 2016.

RESEARCH & WORK EXPERIENCE

Senior Data Scientist — Mar 2022 – Present
Mastercard — Vancouver, BC

Software Developer/Data Scientist — May 2017 – Feb 2022
Fortinet — Burnaby, BC

Software Developer (Co-op with UNB) — Jan – Aug 2015
IBM — Fredericton, NB

SELECTED PROJECTS

Bridge Hand Recognizer — 2022, Personal

Playing card detector for contract bridge based on YOLO [GitHub]

- Prepared training data from scratch by taking card photos, programmatically combining with different backgrounds, and using OpenCV to virtually adjust lighting, etc.
- Refined and re-sampled training data based on discrepancies of evaluation results among classes to improve performance iteratively.

- Trained models with cloud resources to recognize 52 classes with 99% precision.

Labeling Frauds with Limited Ground Truth 2022, Mastercard

Label generator for optimizing weights of anomaly scores for passive biometrics identity model

- Proposed, experimented, and documented strategies involving data extraction, cleaning, and labeling.
- Ran controlled experiments to identify most influential factors for model performance.
- Filed patent on new labeling methods.

Ensemble Classifier for Malware Detection 2021, Fortinet

Accuracy-focused model built on features from multiple engineering teams, achieving reduced dependencies and better data interaction

- Identified & corrected data quality issue by visualizing hundreds of binary features during EDA.
- Prevented data leakage by consulting data source owners and building solid validation pipeline.
- Stacked on less correlated predictions from linear, tree-based, and deep learning models.
- Achieved at least 95% performance of rule-based team of 30+ malware analysts.

Anomaly Detection in Network Traffic 2019, Fortinet

Open-ended project to discover best approaches to detect abnormal patterns in second-level network traffics

- Applied feature engineering tricks before running clustering algorithms to deliver instant values.
- Further built probability distribution-based detector according to academic papers.

Machine Learning for Wind Power Predictions 2016, UNB

Empirical study of nine machine learning regression models for predicting wind energy

- Collected and merged data from 3 distinct sources to prepare 7 datasets for training & evaluation.
- Designed and conducted experiments to evaluate effects of different structures of DNNs.
- Applied cross-validation and statistical test in order to compare model performance.
- Concluded by visualizing and bench-marking that SVM was best of nine.

Image Segmentation with Parallel k-means 2014, UNB

Project for graduate course Parallel and Distributed Processing

- Parallelized k-means clustering with OpenMP for image files in C/C++ according to past research.
- Compared results for different iterations, k's, and local minima.
- Wrote formal report in format of IEEE Journals.

OTHER RELEVANT R&D PROJECTS

Ranking File Execution Info with TF-IDF Techniques 2019, Fortinet
Finding Related Malwares with Graph Database and Cosine Similarity 2018, Fortinet
Malware Segmentation with Text Clustering 2018, Fortinet
Titanic Data Visualization [Visualization | Code] 2016, Udacity
Research Report on GIB: an Adversarial Search Algorithm for Contract Bridge 2014, UNB
Interactive Rubik's Cube Simulation [Cube | Recorded presentation] 2014, UNB

RELEVANT SKILLS AND TRAINING

Programming Languages	Python, SQL, JavaScript, Java, C/C++
Frameworks & Libraries	pandas, scikit-learn, NumPy, SciPy, Matplotlib, OpenCV, imgaug Spark, TensorFlow,
Other	git, Bash

OTHER TRAINING

Image Processing with Python Track 2020, DataCamp
Data Analyst Nanodegree 2017, Udacity
Time Management 2015, Mitacs

MISC INFO

Volunteering: Recommendation system for Canoo (ICC) • Mastercard Changeworks (CanadaHelps)
Languages: English, full professional • Chinese Mandarin, native
Hobbies: Photography (instagram.com/liuyq983/) • Contract bridge