

Yang Lin Ph. D.

Email: linyanguyc@gmail.com

Phone: (917)-930-7106

Linkedin: <http://www.linkedin.com/in/yang-lin-ny>

Github: <https://fjxmlinayang.github.io/>

SUMMARY

- 10+ years of **research experience** in the area including optimization, statistics, machine learning and risk
- 4+ years in **applying machine learning** (ML) models in technology, finance/insurance, IoT/ energy, and healthcare areas
- 2+ years in **project experience** on machine learning, web development, and mobile development
- Significant ability in communication, teaching, and work independently or as a part of a team

PROFESSIONAL EXPERIENCE

Research Scientist

Dec 2021 - Present

Munich Re/ HSB

- Collaborated with clients and colleagues in deal teams to assess, model, and underwrite the performance risks of ML systems
- Lead the risk assessment for evaluating artificial intelligence product system design and ML system performance for start-ups company (including application areas: financial fraud /IoT sensor detection, etc)
- Designed the customized insurance product and data-driven price model to offer performance guarantee solution for AI product
- Lead publication-oriented research on distribution-free uncertainty quantification application in risk and ML development
- Sourced business opportunities with AI start-ups and the IoT industry and presented deal opportunities to senior management

Adjunct Professor/ Postdoc Researcher

Aug 2020 -Dec 2021

Stevens Institute of Technology, Hoboken, NJ

- Designed hybrid optimization and ML models for maximizing the value of pumped-storage hydroelectricity(PSH) resources
- Implemented reinforcement learning models (approximate dynamic programming) by python to obtain the arbitrage policy and increased the profit by 11%
- Visualized data to analyze the energy dispatch for developing reliability and efficiency market planning and operations and searched features and performed variable analysis and selection to model water storage
- Applied parallel computing to speed up the offline learning algorithm by 10 times in python multiprocessing module
- Initiated the first optimization application course in data science and taught more than 50 undergraduate students.

Doctorial Researcher

Sep 2014 -Aug 2020

Stevens Institute of Technology, Hoboken, NJ

- Researched on statistical inference of stochastic optimization problems in quantitative risk management and machine learning
- Created kernel smooth model by nonparametric statistical methods to approximate risk-neutral/averse composite stochastic optimization problems
- Established the consistency, convergence, stability, and bias reduction of the smooth estimators and extended the method to regression models, classification problems, and optimization problems with financial risk metric
- Reduced mean square error(MSE) of data-driven optimization model on the average value at risk(AVaR) model by 16.7% and higher-order measure of risk by 25%

SKILLS

Programming Languages: Python, Java, JavaScript, SQL, C++

Data Science:

Machine Learning Models: Quantile Regression, Boosting Tree, LSTM, Clustering, PCA, etc;

A/B Testing, Experimental Design, Hypothesis Testing, Exploratory Data Analysis, etc;

Big Data Engineering/ Database: MySQL, Spark

Framework and Platforms: Spring, Hibernate, React, AWS, EC2, GCE, GCS, etc;

Web Development: Node.js, HTML & CSS, React

EDUCATION

Ph. D. in Applied Mathematics

May 2020

Stevens Institute of Technology, Hoboken, NJ

Dissertation: Kernel Smoothing in Sample-Based Optimization

M.S. in Mathematics

May 2014

New York University, New York, NY

B.S. in Mathematics, Information and Computational Sciences

May 2011

Xiamen University, Fujian, China

PROJECTS

Topic includes business index prediction, recommendation engine system, semantic analysis, image recognition

Semantic Analysis for Youtube User Comments Dataset

- Built a machine learning model to analyze and classify users based on their Youtube videos comments on cats and dogs
- Extracted and transformed users' comments via RegexTokenizer and Word2Vec with SparkML pipeline
- Trained logistic regression, Random Forest, and GBT models, tuned hyperparameters, and selected the best model(AUC=0.91)
- Extract insights about cat and dog owners and identified creators with cat and dog owners in the audience

Natural Language Processing(NLP) and Topic Modeling on User Review Dataset

- Discovered and studied main topics of an E-Commerce business customer review dataset for watch by NLP
- Processed review texts to extract features by tokenization and stemming
- Constructed the term frequency-inverse document frequency(TFIDF) matrix to build the unsupervised learning model
- Trained model including K-Means and Latent Dirichlet Allocation(LDA) to extract the main topics on watch review(good, nice, not cheap, etc.)

Movie Recommendation Engine Development

- Developed data ETL(extract, transform, load) pipeline to analyze movie rating dataset with Spark SQL
- Implemented the Alternative Least Square algorithm, tuned hyper-parameters with Spark ML toolbox, and monitored data processing performance via Spark UI on AWS(RMSE=0.7)
- Provided personalized movie recommendations and developed user-based approaches to handle system cold-start problems

Car Image Recognition and Classification based on Deep Learning

- Built a deep learning model to classify the car image
- Trained Convolutional Neural Networks (CNN) model by utilizing the pre-trained model and fine-tune on current dataset
- Deployed the built transfer learning model and predicted the car's name on single images (Accuracy=0.95)

Stock Prices and Market Index Prediction

- Analyzed the volatility feature of stock price and market index based on Pytorch
- Built a deep learning time series model (Long Term Short Memory) by 7 days in advance to predict the stock price
- Trained LSTM model by changing activation and regularization function on GPU
- Deployed the built LSTM model as a service to predict the variation of the S&P 500 index

Customer Churn Prediction in Finance Industry

- Built algorithms for the bank to predict customer leaving based on information and performance records
- Trained supervised learning models including Naive Bayes, Logistic Regression, Random Forest(RF), Gradient Boosting tree(GBT), K-Nearest Neighbors, applied regularization with optimal parameters to overcome overfitting
- Evaluated model performance (Accuracy 0.86) via k-fold cross-validation technique and analyzed feature importance to identify the top factors that influenced the results

PUBLICATIONS

- Darinka Dentcheva, Yang Lin, Spiridon Penev. Stability and Sample-Based Approximations of Composite Stochastic Optimization Problems; *accepted by Operations Research*, 2022
- Yikui Liu, Yang Lin, Lei Wu, etc. Optimal SOC Headroom for Pump Storage Hydropower Dispatched by ISOs: Maximizing Revenue in Day-ahead and Real-time Markets; *under review by IEEE Transactions on Sustainable Energy*, 2022
- Darinka Dentcheva, Yang Lin. Bias Reduction in Sample-Based Optimization; *SIAM Journal on Optimization*, 2022
- Yang Lin. Kernel Smoothing in Sample-based Optimization; *ProQuest Dissertations Publishing*, 2020
- Darinka Dentcheva, Yang Lin. Kernel Estimation of Composite Risk Measures and Asymptotic behavior of Kernel-Based Optimization Problems; *in preparation*

INVITATION TALKS

- Insuring Learning Models: Applied Uncertainty Quantification. *New England Statistics Symposium(NESS)*, University of Connecticut, CT, 06269
- Kernel Smoothing in Sample-based Optimization. *Doctoral Student Celebration*, Stevens Institute of Technology, NJ, May 2020
- Bias Reduction in Stochastic Optimization. *Student Seminar Research*, Stevens Institute of Technology, NJ, September 2019
- Kernel Estimation in Stochastic Optimization with Composite Risk Functionals. *The XV International Conference on Stochastic Programming (ICSP 2019)*, NTNU, Norway, July 2019

AWARDS

US Junior Oberwolfach Fellows

Excellence Doctoral Fellowship

Math Summer Research Scholarship

Mathematisches Forschungsinstitut Oberwolfach, 2018

Stevens Institute of Technology, 2019-2020

Stevens Institute of Technology, 2019