MOHAN LIU

628 Library Place, B3N, Evanston, IL, 60201
mohan@u.northwestern.edu
+1 847 644 9861

EDUCATION

Northwestern University (NU), Evanston, IL — Sep. 2013 - June 2019 (Expected)
- PhD candidate, Department of Materials Science and Engineering

Nanjing University (NJU), Nanjing, China — Sep. 2009 - July 2013
- Bachelor of Science in Physics.

University of California, Los Angeles (UCLA), Los Angeles, CA — July 2012 – Sep. 2012
- Exchange Program: Cross-disciplinary Scholars in Science and Technology (CSST)

TECHNICAL SKILLS

Python: 5+ years, including data mining and visualization, web scrapping and machine/deep learning.

Data analysis skills: Proficiency with Pandas, MySQL and TenserFlow; some experience using R and Perl.

Other skills: Expert knowledge with C++, Git, Docker, AWS, web API and HTML.

RESEARCH EXPERIENCE

Department of Materials Science and Engineering, NU, IL — Sep. 2013 - June 2019 (Expected)

Research Assistant, Advisor: Professor Chris Wolverton

- Materials informatics: large databases and machine learning for materials design and discovery
- Managing and maintaining a computational materials database (using MySQL) containing calculated physical and chemical properties for >600,000 compounds.
- Developed a simple and efficient web API based on REpresentational State Transfer (REST) principles to provide the community with an easy access to our database.
- Trained three regression models (*LASSO, SVR and random-forest*) using our materials dataset and predicted materials band gaps with ~20% relative RMSE. Constructed our feature space using elemental-property-based attributes and perform univariate feature selection to reduce feature dimensions.

DATA SCIENCE EXPERIENCE

- Home Credit Default Risk: predict how capable each applicant is of repaying a loan — May 2018
- Applied one-hot encoding for categorical features and preformed PCA to reduce dimensions. Conducted recursively feature elimination to further extract the most important features.
- Trained a large dataset with more than 300,000 observations using *gradient boost decision tree model* (using LightGBM) and our predicted area under the ROC curve is 0.796.
- Ensemble and cross validation have been utilized to overcome potential overfitting. Amazon Web Services (AWS) has been used to accelerate training process.

- Avito Demand Prediction Challenge: predict demand for an online classified advertisement — June 2018
- Applied latent semantic analysis for natural language processing (using NLTK and polyglot) to extract test features from advertisement titles and descriptions. Trained *convolutional neural network* (using TenserFlow) to collect image features from images of items provided by each seller.
- Stored large datasets (more than 15 GB in total) using Hierarchical Data Format (HDF) and Feather format to facilitate data I/O. Performed multiprocessing to accelerate feature engineering process.
- Trained the dataset with more than 1,500,000 observations and 900 features using *gradient boost decision tree model* (using LightGBM) and our predicted RMSE is 0.2236 (ranked top 11%).

PUBLICATIONS & PRESENTATIONS

M. Liu, W. F. Schneider and C. Wolverton, *Configuration-dependent adsorption energy modifications at bimetallic surfaces,* NAM, oral presentation, June 4-9, 2017, Denver

L. Huang, P. C. Chen, M. Liu, X. Fu, Y. Yu, P. Gordiichuk, Z. Xie, C. Wolverton, Y. Kang, C. A. Mirkin, *Catalyst design by DFT and scanning probe block copolymer lithography,* PNAS (2018)