Wang Yao

Phone: (732) 447-6825 E-mail: yaowang74@gmail.com Work Authorization: **U.S. Citizen**

SUMMARY

5+ years experience in data science and machine learning, a well-trained scientist in mathematical optimization and algorithm development, passionate presenter with extensive knowledge in numerical methods for solving large scale statistical learning problems.

EDUCATION

Ph.D. in **Operations Research**, **Rutgers University, New Brunswick, NJ** 2010-2016
B.A. in **Mathematics**, Minor: **Statistics**, **Rutgers University, NJ**. *Magna Cum Laude* 2006-2010

RESEARCH

Ph.D. Dissertation: Approximate Versions of Alternating Direction Method of Multipliers

- Developed 3 new numerical algorithms for **multi-block optimization** models in machine learning
- Implemented new engine for common optimization problems in ML with **Matlab**, wrote a library including **conjugate gradient, L-BFGS** and **gradient decent** algorithms as sub-solvers
- Overall computing overhead is reduced by 25% compared with classical frameworks

AWARDS

Honeywell Bravo Employee Recognition Award June 2017

WORK EXPERIENCE

Honeywell International Inc. **Data Scientist, Data Science CoE** Oct. 2016 - Present

- Employed business goals to guide data science practice to win contracts, engaged with the clients, data scientists and implementation team in presenting, mentoring and advising
- Led data engineers in India to clean and ingest over 1TB raw power consumption data to **Azure** cluster using **pyspark** and **Hive** database, created reusable library for the team
- Used **R** to implemented **RBF Neural Network** for short-term power load **time series** prediction
- Performed large scale statistical analysis and profiling to segregate population on cluster with **sparklyr** trained power usage model by utilizing external property and weather data
- Built and deployed an energy theft detection model using features from water usage and smart meter events/status data, helped customers to recover at least 3% of their revenue

Chubb Corporation **Claim Actuarial Modeler, Advanced Analytics** May 2015 - Oct. 2016

- Performed data engineering on ∼10 million rows of bodily injury claims/medical data from 7-Eleven
- Created ∼100 numerical, categorical and **text mining** variables, built decision tree severity model for at first contact and logistic regression model at 6-month evaluation point using **SAS EG**
- Conducted **social network analysis** and established network-revenue model for underwriters using **iGraph**, helped them to create 5% more revenue overall by improving network structure
- Created **SAS** codes for monitoring and tracking model performance in production system

Novartis Pharmaceuticals **Intern, Advanced Quantitative Science** May - Aug. 2014

- Learned data science models in pharmaceutical and web development skills using **Linux**, **Apache** web server, **MySQL**, **Python**, **R**, **Javascript** and **HTML5**
- Designed and implemented an interactive web application for clinicians to compare simulation results of dosage selection for different groups of patients, effectively reduced communication overhead

DHS - Rutgers CoE (CCICADA) **Advanced Analytics Researcher** Jan.-Jun. 2013

- Captured and analyzed complex boat allocation data given by U.S. Coast Guard to establish the optimal boat-sharing integer programming model with **Xpress-Mosel**
- Recommended the best boat-sharing plan for U.S. Coast Guard to cover required mission hours under tight budget, presented the analysis and delivered the product to the C-level management

Eli Lilly and Company **Intern, Data Visualization** May - Aug. 2012, and Jul. - Dec. 2013

- Tested extensively the Bioprocess Data Collection System Data Mart(BDCSDM) using **SQL**
- Developed a data visualization tool that is completely database driven and maintenance-free
- Developed automated filtration experiment data acquisition application to interface with previously under-utilized costly advanced **IoT** equipments