Wang Yao

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

SUMMARY

5+ years experiences 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 30% compared with classical frameworks

WORK EXPERIENCE

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

- Lead data engineers to cleanup over 1 TB raw power consumption data consists of 1.5 million text files on **Azure** cluster through creating reusable **pyspark** library and establishing **Hive** database
- Selected features from weather data, day of week and partial ACF, implemented a **RBF Neural Network** model with **R** for short-term power load **time series** prediction
- Utilized external property data, carried out univariate analysis and segregated population into 15 micro segments by maximizing the margin of average power consumption
- Trained unusual usage model on segment level on cluster using **sparklyr** for specifying abnormal observations to convert extremely imbalanced population into balanced training dataset
- Using water usage, smart meter events/status data, trained a energy theft detection model
- Received Bravo recognition award, heavily interacted with implementation team, data scientists and business via advising, mentoring and presenting

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

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

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

- Designed an interactive web application for clinicians to compare numerical simulation results of dosage selection for different groups of patients
- Implemented the web service using **Linux**, **Apache** web server, along with **MySQL**, **Python** and **R** for server-side, **Javascript** and **HTML5** for client-side programming

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

- Captured and analyzed complex information given by U.S. Coast Guard to establish the optimal boat-allocation 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 and 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) with **SQL**
- Automated the visualization tool to be database driven, eliminating maintenance overhead
- Developed automated filtration experiment data acquisition application to interface with previously under-utilized costly advanced **IoT** equipments