SHITAO WANG

Miami, FL
steven.shitaowang@gmail.com | github.com/shitao | linkedin.com/in/shitaowang

Technical Skills

Languages: Python, Matlab, C++, Fortran, Bash, SQL.
Machine Learning: Classification, Regression, Clustering, Q-Learning.
Statistical Analysis: Principal Component Analysis, Uncertainty Quantification, Bayesian Inference, Sensitivity Analysis, Markov Chain Monte Carlo, Ensemble Methods, Karman Filter, Predictability.

Experience

Insight data science – Data Science Fellow **January 2017**

- Built a web application in python to provide skill suggestions for data science career.
- Crawled 3000+ job postings using beautifulsoup and re, cleaned data in pandas and stored in a PostgreSQL database.
- Constructed a topic frequency library for job descriptions and ranked the missing skills based on the topic frequency.
- Deployed the interactive web application using Flask, Bootstrap, Javascript and AWS.

Graduate Student – University of Miami **2012-2017**

Quantifying the Predictability of an Ensemble Oceanic Forecast using Information Theory

- Evaluated the model performance against a large amount of unstructured observations (~10TB).
- Applied information theory to quantify the predictability of an oceanic state forecast.
- Implemented a Bayesian Inference approach to correct the initial condition uncertainty given observations.

Performance Comparison of Different Uncertainty Quantification Techniques

- Compared four different ensemble-based techniques for uncertainty quantification.
- Gauged the performance of these techniques using a Monte Carlo method.
- Applied Gaussian Process and Compressive Sensing type techniques to handle the noise in the data.
- Performed sensitivity analysis to identify the principal contributor to the model uncertainty.

Side projects

Kaggle: House Prices Advanced Regression Techniques -*github.com/Shitao/udacity_capstone*

- Applied ensemble learning technique to predict the house price using 79 features, ranked top 10 (7/632) at the time of submission.

Build a web application for stock price prediction - *shitao-stock.herokuapp.com*

- Applied gradient boosting tree models to predict the stock price for the next 7 days using Wiki EOD Stock Price data in real time.

Train a Smartcab How to Drive - *github.com/Shitao/smartcab*

- Applied reinforcement learning, specifically Q-Learning algorithm, to build a simulated navigation agent.

Creating Customer Segments - *github.com/Shitao/customer_segments*

- Built a K-Means clustering model to create customer segments in order to optimize the operation resources.

Education

University of Miami – Marine Physics, Ph.D. **Summer 2017 (expected)**
Udacity – *Machine Learning Engineer Nanodegree* **October 2016**
University of Miami – Marine Physics, M.S. **May 2015**
Ocean university of China – Marine Technology, B.S. **May 2012**