

Zihao Wang

zw1074@nyu.edu | (551)225-9955 | 204 10th St, Apt. 314, NJ 07302 | <https://zw1074.github.io>

Educations

New York University, Courant Institute of Mathematics

New York, NY

Master of Science in Data Science (3.80/4.0)

May 2017

Nanjing University

Nanjing, China

Bachelor of Science in Computational Mathematics (3.50/4.0)

Jun 2015

Related Courses

Machine Learning, Big Data, Deep Learning, Statistical Inference, NLP, Advanced Python, Fundamental Algorithm, Business Understanding for Data Science, Artificial Intelligence, Operating System, Optimization Theory, Linear Algebra, Probability Theory, Stochastic Processes.

Skills

- Programming Skills: Python (Proficient), C/C++ (Proficient), SQL, Hadoop, AWS, Git, Theano, TensorFlow, Matlab
- Data Analytic Skills: Machine Learning, NLP, Deep Learning, Web Scrawling, Data Visualization
- Language: English, Chinese

Professional Experience

American International Group Inc.

New York, NY

Deep Learning Research Assistant, Intern

Aug 2016-Dec 2016

- Contribute to the automatic car damage appraisal (ADA) project, especially for license plate detection on poor image dataset and heat map generation of damaged parts.
- Build an efficient license plate detector by fine-tuning a simple but efficient convolution neural network to generate saliency map and OpenCV to extract contour of license plate.
- Build an end-to-end Grad-CAM solution to extract information about damage part in a pre-trained model and image.
- Implement these methods with Theano and Tensorflow and test them on both Linux and Windows system.
- The license plate detector achieves 15% top-5 error with 30s/image speed.
- The map generation API has been merged into ADA project and proven to be more effective than previously used.

Academic Projects

Center for Data Science, NYU

New York, NY

Deep Learning: Real Time Data Ingestion and Anomaly Detection for Particle Physics

Sep 2016-Dec 2016

- Build an auto-encoder for compressing 100GB CERN particle collision event.
- Evaluate model by calculating R2-score of reconstruction vector and MLP auto-encoder has 0.95 after adding threshold RELU.
- Apply well-trained auto-encoder to create an efficient anomaly detector for detecting abnormal.
- This work is supervised by Prof. Kyle Cranmer and presented in NIPS 2016 invited talk.

Big Data: Explore Relationship Between Citi-bike and weather

Feb 2016-May 2016

- Extract insight relationship between Citi-bike and weather in New York City by using data in 2015.
- Filter and select different features from data concurrently by using Hadoop MapReduce technology in NYU HPS cluster.
- Visualize data by using different technologies such as google map API (gmaps).
- The conclusion agrees with our hypothesis – when the weather gets better, the duration of trips increases.

Machine Learning: Duplication Detection

Feb 2016-May 2016

- Build an end-to-end solution to detect duplication record in health care information system.
- Use 32-bit rolling hash to compute the representation value of string quickly.
- Create our own parallel filter algorithm to find possible duplication pairs which is much faster and more effective than local sensitive hashing (LSH) algorithm.
- Use T-SNE technique to visualize the feature vector and train a random forest model based on these vectors.

Machine Learning: Yelp Restaurant Rating Prediction

Feb 2016-May 2016

- Build a model to predict future rating level of restaurant based on business attributes, previous ratings and Yelp reviews.
- Divide the rating to 3 different levels based on the distribution to make a balance dataset.
- Use Google pre-trained word2vec model to represent Yelp reviews as 300-dimension vector.
- Select models and corresponding hyper-parameters by using cross-validation and AUC-scores.
- The best model is a logistic regression with L1 regularization and its average cross-validation AUC-scores is 0.86.