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 **Shanghai, China**
*Bachelor of Science in **Computational Mathematics (3.50/4.0)*** *Jun 2015*

Related Courses
Machine Learning, Big Data, Deep Learning, 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(Mandarin)

Professional Experience

American International Group Inc. **New York, NY**
Deep Learning Research Assistant, Intern *Aug 2016-Dec 2016*

- Contribute to the whole automatic car damage appraisal project, especially for license plate detection and heat map generation of damage part
- Build a 30ms/Image license plate detector by using convolution neural network to generate saliency map and OpenCV to extract contour of license plate
- Build an end-to-end Grad-CAM solution to generate heat map of damage part given a pre-trained model and image
- Design a heat map generator module for efficiently use and easy updating
- Implement these methods by using Theano and Tensorflow and test them on both Linux and Windows system

Academic Projects

Center for Data Science, NYU **New York, NY**

Deep Learning: Efficient auto-encoder for physics particle collision event *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
- This work has been praised by Prof. Kyle and presented in NIPS 2016

NLP: Image caption *Sep 2016-Dec 2016*

- Build an end-to-end attention-based image caption generator based on MS COCO dataset
- Modify original model by calculating feature map attention instead of location attention
- Gain less than 20 perplexity by using google pre-trained GloVe word representation

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 quick compute the representation value of string
- Create our own parallel filter algorithm to find possibly duplication pair 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 feature rating level of restaurant based on business attribute, previous rating and Yelp reviews
- Divide to rating to 3 different level based on the distribution
- Apply dummy variable on category feature and fill NaN by mean
- Use google pre-trained word2vec model to represent Yelp reviews as 300-dimension vector
- Train a 0.86 micro-AUC score fine-tune logistic regression model by combining above feature