Guangyao Yang

+86 18639095151 • guangyaoyang@foxmail.com
www.younggy.com

Passionate about machine learning and data science, with strong learning enthusiasm and ability, having finished 43 MOOC courses related to machine learning and data science.

Education

- **Huazhong University of Science and Technology** — **Wuhan,China**
  Master of Engineering, Machine Learning — *Sep 2015–Present*
- **Tianjin University** — **Tianjin,China**
  Bachelor of Engineering, Electrical and Electronic Engineering — *Sep 2011– Jul 2015*

Journal Publications

As the first student author, I have published **two SCI papers, one EI paper (General Assembly Excellent Paper)** in the field of partial discharge pattern recognition.

- *Optimal Feature Selection for Partial Discharge Recognition of Cable Systems Based on the Random Forest Method*
- *Rough Set Theory Applied to Pattern Recognition of Partial Discharge in Noise Affected Cable Data*
- *SDMF based Interference Rejection and PD Interpretation for Simulated Defects in HV Cable Diagnostics*

Internship

- **Toutiao (Bytedance)**
  Algorithm engineer on news recommendation — *May 2017–Jul 2017*

  Responsible for news **personalized recommendation**. Used **relevance feedback** to add new tags to news. Generated feedback tags using **SVD, LSTM, multi-head attention**(based on Google's paper). Model was put into use with **ctr increased by 1.1%**.
- **Artificial Intelligence Institute at Sinovation Ventures**
  Algorithm engineer on deep learning — *Jul 2017–Present*

  Developing artificial intelligent Fight the Landlord (a Chinese card game)

Projects

- **Projects of Udacity Deep Learning Nano Degree**

  This is an on-line course teaching deep learning techniques. I've finished all projects. Some projects in the course are: Artistic Style Transfer; Generate new scripts using **LSTM**; Implement **Word2Vec**; Train **seq2seq** for English to French translation; Implement a **deep Q-learning network** to play Cart-Pole game; Use a **DCGAN** to generate house number and human faces images.
- **Projects of Stanford CS231(*Convolutional Neural Networks for Visual Recognition*) Course**

  CS231 is the a course teaching deep learning at Stanford University. I've finished all assignments. Some of the assignments are: Implement **CNN, RNN, LSTM** forward and backward process; Caption images with **CNN** and **LSTM**; Compute **image gradients** to produce saliency maps, fooling images and deep dream.

- **Projects of Udacity Self Driving Car Nano Degree**

  This is an on-line course teaching self-driving car techniques. I've finished project about finding lane lines on the road using **canny edge detector** and **hough transform**.

- **Classification Research of Partial Discharge**

  Solved 5 types of classification problems using **logistic regression, neural network, SVM, and RF**. RF had the highest accuracy of 89.2%.

- **Feature Engineering and Recognition of Partial Discharge**

  Constructed time-domain and frequency-domain characteristics; Implemented feature selection using **RF, LASSO**; Implemented feature extraction using **PCA, RBF kernel**; Trained model using **RF, GBDT** with highest accuracy of GBDT 91.3%.

- **Documents Clustering and Keyword Association**

  Clustered documents using **k-means++, GMM, LDA**. Retrieved Documents using **KD-Tree, LSH**. Used the words with the higher scores within the subject in LDA as the association keywords.

- **Analysis of Commodity Communication Network Based on Agent-based Model**

  Constructed small world network to simulate the process of commodity communication using **igraph** and **ABM**. Studied the effects of word of mouth, advertising, promotion strategy on commodity communication. Developed R package.

Machine Learning Knowledge

- **Certification Courses:** 43 course certifications including: Stanford "Machine Learning", "Mining Massive Dataset"; UW "Machine Learning Specialization"; MIT "The Analytics Edge"; NTU "Machine Learning Foundations", "Machine Learning Techniques".
- **Theories:** VC dimension; Regression and Classification(SVM, logistic regression, RF, GBDT); Clustering(k-means, GMM, LDA); Deep Learning(CNN, RNN, LSTM, GAN, Deep Q-learning); Feature selection(RF, LASSO); Feature extraction(Kernel, Auto-encoder, NN);Recommendation(Collaborative Filtering, Matrix Decomposition); NLP(w2c, LSA, seq2seq)
- **Tools:** Hadoop, Spark, Hive, Tensorflow, scikit-learn, pandas, numpy

More about me

- **Blog:** http://blog.younggy.com
- **Github:** https://github.com/YoungGer
- **Courses Certifications:** http://younggy.com/moocs.html