

HAN TANG

16 Victoria Ave, Donnybrook, Dublin 4. D04 T2P1

(+353)0830477054 ◊ D16129273@mytudublin.ie

LinkedIn Profile: Han Tang

Stamp 1G

EDUCATION TRAINING

Technological University of Dublin, Dublin, Ireland

September 2018 - February 2020

MSc in Computing (Data Analytics)

School of Computing

Award: 2.1

Relevant Subjects:

Working With Data (A): Programming in SQL and R.

Machine Learning (A⁻), Deep Learning (A): Machine Learning in Python.

Data Mining (B⁺): The full life cycle of Data Mining project deployment.

Dissertation:

A Comparison Study on State-of-the-art Minority Class Data Oversampling Techniques for Imbalanced Learning

Dublin Institute of Technology, Kevin St., Dublin, Ireland

September 2017 - June 2018

Pre-master for MSc in computing.

Average score: 78

Beijing University of Chemical Technology, Beijing, China

September 2013 - June 2017

BSc in Applied Chemistry

Dissertation:

Study the factors of the layer heights of Layered Double Hydroxides.

PROJECTS

Dissertation: A Comparison Study on State-of-the-art Minority Class Data Oversampling Techniques for Imbalanced Learning

Github Link: Msc Dissertation Han Tang

Study the state-of-the-art approaches to imbalanced learning. The focus of study is on Synthetic Minority Oversampling Technique (SMOTE) and its extensions.

Compare more than ten data resampling methods' performances on 35 datasets, and check whether their differences are in a statistic level.

Propose a data security system for the German Government

Group work: Analyse the security risks in the German government Data Management System and propose measures to minimize them.

As an individual: Establish relevant policies and implementation plans to enforce the security measures.

Multi-label Bird Species Classification - NIPS 2013 Kaggle Competition

Github Link: NIPS 2013 bird song classification

Produce Deep Learning Neural Networks to identify the bird species from their records, trained from a dataset contains 1000 instances and 87 classes.

Plot the Mel-frequency spectrograms of the soundtracks, then to train convolutional neural networks from the spectrogram graphs.

The architecture is from the ConvNet in the package Fastai. The result of AUC is 0.83.

WRITINGS

Effective feature engineering techniques on data set contains sequential feature

Github Link: Ford Challenge

The data set retrieved from Kaggle competition - Stay Alert! The Ford Challenge records the change of drivers' behaviours over time.

Rolling means and standard deviations of each feature are introduced as new features to record the sequential change.

This feature engineering technique improves the performances of models of several machine learning algorithms, proved to be useful.

CAREER EXPERIENCE

Postal Savings Bank of China

July 2019 - September 2019

Data Mining Analyst Intern.

- Conduct data mining, data modelling, statistical analysis. Data analytics support for customer credit risk assessment.
- Arrange, edit and archive user guidance for frequently used machine learning packages such as Scikit-learn, Numpy, Pandas, TensorFlow, and Keras.

State Key Laboratory of Chemical Resource Engineering

November 2015 - May 2017

Researcher

- Research on methods of calculating/estimating chemical parameters of complex compounds.
- Calculate chemical parameters of compounds automatically in Python.

LANGUAGES & SKILLS

Technical Skills

SQL, MS Excel, R
Python(4 years), Machine Learning
Deep Learning, Data Mining
LaTeX, Tableau, SAS
Keras, Pandas, Numpy
Scikit-learn, TensorFlow
Web Scraping

Soft Skills

Analytical Skills, Productivity
Problem Solving, Teamwork
Presentation Skills, Integrity
Critical Thinking, Creativity

Languages

English (Full professional proficiency), Mandarin (Native proficiency)

INTERESTS

Associations:

Member of UK Oracle User Group
Github Page: <http://github.com/HirahTang>.

Other activities:

Play in a weekend football amateur league.
Enthusiastic in mathematics and general science.