HAN TANG

1 Chengnan West Rd., Tianxin District, Changsha, Hunan. 410015
(+86)13520477862 ⋄ D16129273@mytudublin.ie
LinkedIn Profile: Han Tang

EDUCATION TRAINING

Technological University of Dublin, Dublin, Ireland *September 2018 - Febraury 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.

NOTABLE PROJECTS

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

Github Link: Msc Dissertation Han Tang

- This research was carried out using 35 imbalanced datasets, to produce statistically significant conclusions. The imbalanced ratios of selected datasets range from 8.6:1 to 130:1. Datasets for testing include both unstructured and structured datasets. Simple feature engineering and data preparation also conducted.
- Study the state-of-the-art approaches to imbalanced learning. The focus of this research is to compare techniques for imbalanced learning on the data level. Eleven data re-sampling algorithms studied and amended to implement for testing.
- The focus is on Synthetic Minority Oversampling TEchnique (SMOTE) and its extensions. Its extensions are classified to Range restricted and Clustering Based.
- Data re-sampling algorithms implemented using Python. Applied clustering techniques are from framework Scikit-learn.
- Selected predictive modelling algorithms are Decision Tree Classifier, k-Nearest Neighbour Classifier, and Support Vector Machine Classifier.

- Predictive model performances varied by the selected classifiers and data re-sampling techniques, either measured by F1 or by AUC.
- Results evaluation and analyse investigates how machine learning algorithms coordinate with data re-sampling techniques and gives recommendations based on the distribution pattern and types of imbalanced datasets. The differences in performances are to pass statistic tests.

Statistical Analysis and Interference Using R

Github Link: Statisitical Analysis Using R

- The objective of this project was to find the factors that are influential in causing sleep problems to a person.
- A dataset containing information about lifestyle choices, mental health and physical condition of a person was used to determine if any of these factors are related to a person's sleep problems.
- The data was collected through a questionnaire from the staff members of a University in Melbourne, Australia.
- The dataset was explored and a statistical description of the features was produced which included a normality test for distribution of values for each feature, range of values for each feature and identifying outliers.
- A visual representation of distribution of values for the features was produced along with the degree of correlation between each individual feature and the target variable.
- A logistic regression model was trained and evaluated using the categories of features mentioned above.
- Multilevel regression models were trained by adding features progressively at each stage to enhance the feature set. This improved the prediction accuracy of the model at each stage.
- R Markdown format was used to document the implementation details, results and the conclusion of the project.

Portuguese Bank Data - Data Mining Project using SAS Enterprise Miner

- The objective of this project was to identify the customers have a tendency to subscribe to a term deposit account, based on their characteristics and their records of previous campaigns.
- This predictive modelling task was conducted using SAS Enterprise Miner, from feature engineering and data preparation to modelling and evaluating.
- The correlations between each feature and the target variable are investigated. Then models of multiple machine learning algorithms are trained and compared, include Decision Tree, Logistic Regression, and Neural Networks.
- Predictive analyses produce a result of higher than 90% classification accuracy.
- The same process also performed using SQL, under the environment Oracle Schema.

CAREER EXPERIENCE

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.

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.

LANGUAGES & SKILLS

Technical Skills	SQL, MS Excel, R
	Python(4 years), Machine Learning
	LaTex, Tableau, SAS
	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.