

Qirun Chen

<http://oopsryan.github.io>

Email : qirun.chen@ucdconnect.ie

Mobile : 353-087-0937260

EDUCATION

- **University College Dublin**

Dublin, Ireland

Master of Science in Computer Science; GPA: 3.48

Sep. 2017 – Sep. 2018

- **Main Modules:** Data Mining (Rapid Miner), Machine Learning (Weka), Text Analytics, Information Visualisation (D3.js and Tableau), Advanced Machine Learning (scikit-learn, Pandas, and Keras), Connectionist Computing (ANNs), Big Data Programming (bash, Hadoop MapReduce, and Spark), Advanced Data Structures in Java (Algorithms), Recommender Systems & Collective Intelligence (Collaborative Filtering and Crowdsourcing)

- **Ningbo University**

Ningbo, China

Bachelor of Engineering in Software Engineering; GPA: 3.84

Sep. 2012 – Jun. 2016

EXPERIENCE

- **Zalando SE**

Dublin, Ireland

Data Science Intern

Summer 2018

- **Customer Segmentation:** Using Machine Learning techniques to classify customers into one of pre-defined preference types according to their browse data on Zalando online store. This can help identify customers' affinity with different brands, so that Zalando can advertise them similar brands.
- **Dev Environment:** Analytics environment based on AWS, standardized Apache Spark and Scala dependencies. Deployed Spark applications on clusters in Amazon EMR, interacted with customer data(stored in Apache Avro) in S3, and created steps to automatically run pipelines.
- **Modelling:** Performed data cleaning and transformation to create features. Also set thresholds to reduce noise further during label generation. Built a multinomial model(based on Logistic Regression Model in Apache Spark Machine Learning library). Trained the model on last several months of customer browse data, and tuned hyper-parameters using CrossValidation and GridSearch.
- **Evaluation:** Applied basic evaluation metrics on each target preference type. Used lift analysis on top N customers and visualised the lifts to compare with the benchmark.
- **Job Monitoring:** Used Ganglia, a scalable distributed monitoring system for clusters, to monitor the conditions of each worker node, like free memory percentile and CPU IO wait. Inspected Spark job executions in Spark UI.

- **Uni-tech IT Co., Ltd**

Ningbo, China

Software Engineer Intern

Sep 2015 - Mar 2016

- **Web application - Java EE:** Implemented the back-end system for supervisors to assign and ensure officers to supervise tasks, manage information on the app, and analyze data presented by charts.
- **Android Application - Android SDKs:** Made several flexible Android UI by reusing the Fragments in different layouts to enhance user experience, and fixed some display bugs of coordinates on the map.

PROJECTS

- **Data visualisation - Channelling Hans (See Demo) :** Recreating the GapMinder World visualisation using d3.js. A bubble plot representing the countries of the world. The population of each country mapped to bubble area. These countries are described by GDP and Life Expectancy mapped to x and y axis position. Applied appropriate use of colour, and showed the ability to view data for a particular year and animate the change in country statistics from year to year 1900 – 2016.
- **Super Learner Classifier - Stacked Ensemble Algorithm (See Demo) :** Implementing the stacked ensemble classifier described in (van der Laan et al, 2007) based on scikit-learn. This is a classification model that uses a set of base classifiers of different types, the output of which are combined in another classifier at the stacked layer.
- **Multi-layer Perceptron:** Implementing the implementing a simple multi-layer perceptron using Backpropagation algorithm with stochastic gradient descent. It can learn XOR and Sin functions very well. Trained the model on public Letter Recognition Data Set(clean and extracted features) to recognise people's handwriting.