
Yuyu Liao

目前位於果實夥伴 (OneAD) 擔任資料分析師 (Data Analyst) 一職，在資料科學領域有4年經驗，使用語言主要為 Python 和 R。熱衷於資料分析、機器學習與深度學習等技術鑽研，並透過理解理論與實作的過程，將相關研究的內容用來解決資料產品上的各種問題。除此之外，樂於探究資料工程方面的各種工具的使用與系統架構上的設計，朝著資料科學的道路上不斷的學習與拓展新知。



roymond20@gmail.com, 0920-117-981, Taipei, Taiwan

Skills

Machine Learning

- Supervised learning
 - Naive Bayes, Random Forest, AdaBoost
 - GDBT, XGBoost, LightGBM
- Unsupervised learning
 - K-mean
- Dimension Reduction
 - PCA, LDA, t-SNE

Deep Learning

- Framework: Tensorflow, Keras, Pytorch
- Research Field:
 - Image classification
 - Object detection
 - Natural Language Processing (NLP)

Tools

- Apache Spark
- Apache Airflow
- Apache Superset
- MySQL
- Google Cloud Platform (GCP)

Language

- Python
- R

Data Visualization

- R-Shiny, ggplot2, plotly

Experience

Data analyst at OneAD, Mar 2017 - Present

1. Data ETL 流程
使用 Spark 與 Airflow 等相關工具建立完整的 Extract-Load-Transform 的資料流程，加速團隊在資料的使用上更有效率。
2. 跨裝置配對 (Cross Device Matching)
根據 probabilistic of cross device tracking 的定義，利用使用者觀看廣告的相關數據 (Ex: title / url / IP / time ...etc)，透過 word2vec / pair-wise similarity computation / session ID 等方法，來處理在 ID tracking 中用戶跨裝置的問題。
3. 使用者留存率
定義適合廣告產業的留存公式，分為 sequence 與 commoms 兩種觀察使用者的留存率的方式，追蹤使用者留存的變化，同時也成為其他部門在優化 ID tracking 部分時的重要參考指標。

4. 廣告推薦系統

使用 **Tensorflow** 框架實作 [Training Deep AutoEncoders for Collaborative Filtering](#) 與 [Collaborative Denoising Auto-Encoders for Top-N Recommender Systems](#) 兩篇論文的方法，來建立廣告推薦系統，最終提升廣告觀看率從原本約 20% 的觀看率提升到 40%。

Code : <https://github.com/roymondliao/recommender>

5. 使用者輪廓預測 (Demographic Prediction)

藉由使用者觀看廣告的相關數據，經過 features engineering / XGBoost / LightGBM / ensemble learning 等方法，來預測使用者的年齡與性別。

Data analyst at Juvo+, Mar 2016 - Jan 2017

1. 產品銷售預測系統

由過往產品銷售的數量，使用指數平滑法對產品進行下一期的銷售量預測，提供採購人員在訂購各個產品下一期所需的數量上做為參考指標，並且降低採購人員在採購上發生過度訂購或是過少訂購的狀況，節省了庫存成本。

2. 季節性產品分析

利用指數平滑法與卡方檢定方法，來判斷產品的銷售狀況是否含有季節性的因素，並提供相關數據與圖表給採購人員，以利安排下一季預計採購的產品。

3. 銷售排名與銷售數量相關性分析

銷售排名 (sales rank) 與銷售數量 (sales quantity) 是在亞馬遜平台 (amazon) 上各大賣家所專注研究的數據，也因此，參考了 [The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve](#) 這篇論文，並將論文中所提出的 negative binomial regression model 套用於內部資料上，以內部資料為基準，用來預估其他賣家的產品可能的銷售數量。

Data analyst at Cheetah Mobile, Jul 2014 - Feb 2016

1. 使用者留言自動回覆系統

透過網路爬蟲的技術爬取 Google play 平台上的使用者留言資料，並且與產品經理、客服團隊合作，在初期建立特定關鍵字庫以及結合使用者留言分析系統的結果，再經由內部後台系統的整合，建立自動回覆留言的機制。

2. 使用者留言分析與分類系統

使用 Text mining 的方法來處理使用者在平台上的留言，並將留言標記為 Postive / Negative / Spam 三種類別，使用 naive bayes algorithm 來建立留言分類模型，其目的在幫助客服團隊在處理大量用戶留言資料時只需專注特定類別的留言處理並回覆，大幅節省了過濾使用者留言的時間。

3. 資料視覺化平台

依據使用者留言分析系統所得到的相關數據，再結合各項產品的相關數據，依據時間維度的不同，使用 R Shiny Server 建立視覺化的互動圖表，讓團隊可以輕易地了解用戶對於各項產品的喜好程度，並且追蹤每日、每週與每月用戶的活躍程度，以及呈現其他相關的訊息給團隊參考。

RMA Quality Analyst at GIGA-BYTE TECHNOLOGY CO., LTD., Jul 2013 - Apr 2014

1. 產品維修數據分析

2. 零件備料與預測

3. 顯示卡採礦數據分析與效能調校 (例如:比特幣、萊特幣)

4. 維修經費分析與審核

5. 維修端與業務端事務協調與處理

TalkingData Ad Tracking Fraud Detection Challenge

比賽目的：預測使用者在點擊手機 app 上的廣告後，是否會下載應用程式。

方法：

1. 透過不同維度 (例如：時間、app、裝置、OS..等) 來產生新的 features 進行模型的建模。
2. 建模採用 XGBoost / LightGBM / Catboost 三種演算法來進行模型的建立。
3. 最終將三種演算法預測出來的結果以 ensemble models 的方式作為最終預測的模型。

Airbus Ship Detection Challenge

比賽目的：依據衛星圖像的圖片，希望能找到圖像中船隻出現的位置，藉此來提高檢測船隻是否有出現的準確率。

方法：

1. 先透過建立 classification model 的方式，來判別圖中是否有船隻的出現。
2. 結合 ResNet 與 U-Net 兩種演算法，對圖像判別出來有船隻的圖片，進行船隻位址的檢測。

Reference : <https://www.kaggle.com/yuyuliao/competitions>

Side Project

NBA比賽預測系統

由於本身喜歡打籃球，也愛觀看 NBA 比賽，並且認為運動彩卷是一種投資方式，所以開啟預測比賽的想法。參考 [Basketball-reference](#) 網站上提供的相關數據以及 *Dean Oliver* 所提出的四因子模型 (four-factor model) 的為基礎，建立了差異四因子模型 (difference four-factor model) 來預測賽事的結果。

使用 R 來完成下列工作流程：

1. Web Crawling
2. Data Clean
3. Machine Learning
4. Data Visualization

Code : https://github.com/roymondliao/NBA_Project

參考書籍：

1. Basketball on Paper : Rules and Tools for Performance Analysis
2. Mathletics : How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football

Certifications

Machine Learning, Feb 2017 - Present

<https://www.coursera.org/account/accomplishments/certificate/JSNE5CYQ4FDG>

Data Science: Deep Learning in Python, Feb 2017 - Present

<https://www.udemy.com/certificate/UC-0NYWYTAM/>

Machine Learning A-Z™: Hands-On Python & R In Data Science, Jan 2017

<https://www.udemy.com/certificate/UC-TFHKWIG9/>

EDUCATION

- 統計研究所 2009 - 2011
碩士, 私立逢甲大學, 台中
- 統計系 2004 - 2009
學士, 私立銘傳大學, 桃園