

## 資訊管理期末報告專案提案 - 第八組

組員姓名與學號

R09725008 林聖典

R09725009 顏煥勳

R09725011 陳宇鑫

R09725024 彭琮鈺

R09725052 余俊廷

### 一、研究動機

線上評論在社群平台中是一種熱門的資訊分享形式，使用者能夠依據自身經歷將與店家的消費經驗分享至平台上。在電子商務為消費主流的網路時代，多數消費者在決定是否要購買產品之前會蒐集其他消費者的相關心得評論，若是看到該店家的評論多為正向的將會提升自己的購買意願，反之亦然。然而有些不肖業者為了使自家產品在評論上脫穎而出，便會聘請一些『專業寫手』替該業者撰寫不符合事實的『假評論』，企圖欺騙消費者進而影響消費者的購買意願。

在台灣最知名的便是 2013 年的三星寫手門事件<sup>1</sup>，三星透過公關公司鵬泰操作議題，散播 HTC 的不實使用心得藉此抹黑競爭對手，後來被台灣的公平交易會裁定違法並開罰 1000 萬元。又或者是中國知名的旅遊網站《馬蜂窩》<sup>2</sup> 被媒體踢爆該網站超過 85% 的點評是透過虛假帳號所產生，以上事件皆顯示虛假評論會對於社會存在不良影響。

Yelp 為美國知名的評論社群平台，它不只是美食情報網，其內容也涵蓋購物、食品、美容、運動休閒等類別，使用者可針對特定店家進行評論、上傳照片，或是將評論內容轉發分享他人，以及配合帳號登入進行收藏等操作。然而過去研究指出在 Yelp 平台中有將近 20 % 的評論是由受僱的寫手撰寫出的假評論<sup>3</sup>，Yelp 也因為假冒評論情形過於嚴重，曾在 The New York Times 撰文批評與公開他們所發現的不肖業者<sup>4</sup>。

因此為了提供一個更為可信的評論資訊，本計劃目標在建立出一個假評論分類模型能夠從眾多評論中過濾出虛假評論，協助平台業者 ex. Yelp 過濾虛假評論並重新計算店家評分，使用戶能夠更有參考性的依據評論做出購買決策，讓消費行為不會受到虛假評論而有所影響。

---

<sup>1</sup> <https://www.setn.com/News.aspx?NewsID=2400>

<sup>2</sup> <https://www.fountmedia.io/article/1460>

<sup>3</sup> <https://arxiv.org/abs/1708.08151>

<sup>4</sup> <https://www.nytimes.com/2012/10/18/technology/yelp-tries-to-halt-deceptive-reviews.html>

## 二、文獻探討

假評論預測的研究有使用監督式學習的方式透過語言學特徵 Linguistic feature 以及用戶行為 User-behavioral feature 來建立分類模型，語言學特徵包含 word unigrams、bigrams、語文探索與字詞計算 LIWC、詞性分析 POS 等，用戶行為特徵則涵蓋如平均留言長度、評論星星數標準差等，而目前在機器學習方法中常見的分類模型有四種 Logistic Regression、Naive Bayes、Support Vector Machine 以及 XGBoosts，也有以深度學習方法如 LSTM、CNN 建構分類模型，以下將分別介紹 XGBoost 與 LSTM 的方法：

XGBoost：

XGBoost 是基於 Gradient Boosted Decision Tree (GBDT) 改良與延伸，是一種基於決策樹的集成機器學習算法，採用了梯度提升 (Gradient Boosting) 框架，常被應用於解決監督式學習的問題。在許多資料分析的競賽，例如 Kaggle，都有良好的表現。

長短期記憶模型 LSTM：

與一般的時間序列神經網路相比，長短期記憶模型 (LSTM) 增加 3 個不同的閘門，輸入閘 (Input gate)、遺忘閘 (Forget gate)、輸出閘 (Output gate)。LSTM 透過記憶細胞的狀態 (Cell state)，可以通過門控狀態來控制傳輸狀態，記住需要連續記憶的，忘記不重要的資訊，目前 LSTM 已經經常用於各種建模與預測問題，例如文字生成、機器翻譯、語音識別、生成影像描述和視訊標記等。

除了上述對於機器學習方法的文獻探討，過往也有許多對於虛假評論偵測的模型，以下分別針對兩篇虛假評論偵測論文進行探討：

論文 1：Fake Review Detection on Yelp Dataset Using Classification Techniques in Machine Learning <https://ieeexplore.ieee.org/document/9055644>

論文提出了運用四種不同的機器學習方法建立假評論偵測模型，並且比較各個不同方法間假評論偵測模型的差異，但是論文並沒有使用到深度學習方法，以及對於真實評論與寫手之評論做更進一步的探討分析。

論文 2：應用深度學習技術於網路虛假評論偵測 <http://jeb.cerps.org.tw/files/JEB2019-008.pdf>

論文使用深度神經網路 (DNN)、卷積神經網路 (CNN)、長短期記憶 (LSTM) 等深度學習方法建構網路虛假評論偵測模型，並用於台灣知名論壇的虛假評論上，以提供提供網路評論的平台一個監控的機制，可以預警或是篩選有問題的評論，以避免不實評論對廠商或是消費者造成傷害。

### 三、研究目標與方法

#### 研究目標一：假評論預測模型

本計劃研究目標一即在於將 YelpZip Dataset<sup>5</sup> 所提供的資料集，建立假評論分類模型，並將該模型應用於 Yelp Dataset<sup>6</sup> 中的店家評論進行分類，在過濾出可疑評論之後重新計算店家分數，使用戶能夠更有參考性的依據評論做出購買決策。本研究預計採取數種機器學習方法，以及透過深度學習來建構模型，並且比較各個不同模型間表現的差異。

#### 研究目標二：比較真實評論與寫手之評論差異

本計劃研究目標二在於觀察研究目標一的結果，對標記為虛假的評論與真實的進行比較，檢查假評論是否如預期一樣有較短的篇幅、寫手們會透過特定『模板』來產出假評論、寫手會統一在上班時間(朝九晚五)才會發表評論或是評論是否語意通順，而不是隨意組成，也會將兩者評論做圖表比較，例如透過長條圖、文字雲做進一步分析。

#### 研究目標三：將模型應用於其他平台之可能性 (Google Play, Google Map ...)

本計劃研究目標三在於探討將先前建立的預測模型應用於其他平台上的可能性，不同平台評論的特徵差異可能會對於虛假評論的檢測模型造成影響，但若能夠將模型擴大應用於其他平台上，模型的應用層面將會更為廣泛。

### 四、資料集

本計畫現階段預計使用兩公開資料集，「Yelp Dataset」以及「YelpZip Dataset」，未來可能延伸至自行建構 Google Play 或 Google Map 上的評論資料集作後續相關研究探討。

#### Yelp Dataset

<https://www.kaggle.com/yelp-dataset/yelp-dataset/>

此為 Yelp 官方公開至 Kaggle 平台給有興趣的學術單位作相關研究的資料集，其中包含了橫跨 11 個大都會地區，174,000 間店家資訊以及 5,200,000 的用戶評論。

---

<sup>5</sup> <http://odds.cs.stonybrook.edu/yelpzip-dataset/>

<sup>6</sup> <https://www.kaggle.com/yelp-dataset/yelp-dataset/>

Business		Review		User	
Attribute	Data Type	Attribute	Data Type	Attribute	Data Type
business_id	string	review_id	string	user_id	string
name	string	user_id	string	name	string
address	string	business_id	string	review_count	int
city	string	stars	int	yelping_since	string
state	string	useful	int	useful	int
postal_code	string	funny	int	funny	int
latitude	float	cool	int	cool	int
longitude	float	text	string	elite	string
stars	float	date	string	friends	string
review_count	int			fans	int
is_open	int			average_stars	float
attributes	BusinessAcceptsCreditCards	string		compliment_hot	int
	BikeParking	string		compliment_more	int
	GoodForKids	string		compliment_profile	int
	BusinessParking	string		compliment_cute	int
	ByAppointmentOnly	string		compliment_list	int
categories	RestaurantsPriceRange2	string		compliment_note	int
		string		compliment_plan	int
	Monday	string		compliment_cool	int
	Tuesday	string		compliment_funny	int
	Wednesday	string		compliment_writer	int
hours	Thursday	string		compliment_photos	int
	Friday	string			
	Saturday	string			
	Sunday	string			

[圖一] Yelp Dataest 資料欄位名稱（部分表格）。

## YelpZip Dataset

<http://odds.cs.stonybrook.edu/yelpzip-dataset/>

此為 Prof. Rayana 在 Stony Brook University 做相關研究時所用的資料集，其中涵蓋了在芝加哥地區的餐廳與飯店的資訊，共有由 33,502 個不同用戶提供的 61,541 筆評論，各評論也包含了過濾（假）與否（真）的標籤。

Review metadata:

- date (date)
- review ID (text)
- reviewer ID (text)
- business ID (text)
- label (Y/N)
- useful (integer)
- funny (integer)
- cool (integer)
- stars (integer)
- review (text)

[圖二] YelpZip Dataest 資料欄位名稱。

## 五、預期結果

本研究計畫透過機器學習方法包含SVM、XGBoost、卷積神經網路（CNN）、長短期記憶（LSTM）等方法建構假評論分類模型，模型期望能夠從眾多評論中過濾出虛假評論，協助平台業者 ex. Yelp 過濾虛假評論，使用戶能夠更有參考性的依據評論做出購買決策，讓消費行為不會受到虛假評論而有所影響。

除了建構假評論分類模型，我們也將更進一步比對真實評論與寫手之評論差異，我們列舉了一些可能是真實評論與寫手之評論的差異，例如假評論可能有較短的篇幅或是寫手們會透過特定『模板』來產出假評論等等，我們期望從真實評論與假評論中找出差異，並且進行歸納分析做為未來研究的參考。

最後我們計畫研究將建構好的假評論分類模型用於其他平台之可能性（Google Play, Google Map ...），如果假評論分類模型在其他平台仍然能夠有良好的預測能力，那可以代表我們所建構的假評論分類模型在多個不同的領域 ex. 遊戲、電影都能夠有良好的預測能力，模型的應用層面將會更為廣泛。