

三元決策理論應用於社群媒體 挖掘之情感分析— 以 UBER 在臺營運話題為例

SENTIMENT ANALYSIS BASED ON A THEORY OF THREE-WAY
DECISIONS FROM SOCIAL MEDIA MINING: A CASE STUDY
ON UBER'S OPERATION TOPICS IN TAIWAN

陶冶中 Chi-Chung Tao¹

簡睿志 Rwei-Jhih Jian²

(105 年 10 月 26 日收稿，105 年 12 月 10 日第 1 次修改，105 年 12 月 21 日定稿)

摘要

近年來，「社群媒體」與「共享經濟」結合的行動電商發展方興未艾，其中以 UBER 進軍各國計程車市場所引起的風潮為較知名案例。目前 UBER 在臺灣營運是否合法的爭議不斷，尤其涉及叫車方式、彈性費率、司機乘客互評、稅制等課題，皆面臨適法性之挑戰，因此，若能經由社群媒體挖掘對於 UBER 相關課題之文本進行情感分析，進而了解民眾對 UBER 適法性的情感態度，則可提供政府推動多元化計程車服務之參考。本研究使用爬蟲系統蒐集 UBER 網路文本進行挖掘，並建構三元決策情感分析模式，該模式係將情感傾向區分為正面、中立與負面情感。實證結果顯示，討論聲量前三名的營運話題分別為：營運制度、取締與抗議以及稅務與收費。民眾對於 UBER 在臺營運之網路平台服務的評價，整體情感趨勢係以正面偏中立為主，高於負面情感傾向。整體負面情感趨勢則為 UBER 營運合法

-
1. 淡江大學運輸管理學系副教授（聯絡地址：25137 新北市淡水區英專路 151 號 淡江大學運輸管理學系；電話：(02) 26215656 Ext. 2983; E-mail：cctao@mail.tku.edu.tw）。
 2. 淡江大學運輸科學研究所碩士；現任奕瑞科技公司專案經理。

性之爭議，討論量以非法營業與未盡繳稅義務最多。因此，經由社群媒體挖掘而得出之共同看法是 UBER 應先取得合法執業許可及繳納相關稅費，方可上臺營運。

關鍵詞：三元決策理論；社群媒體挖掘；情感分析；UBER

ABSTRACT

Mobile commerce business combining sharing economy and social media are now in the ascendant, especially UBER launches into taxi markets worldwide which becomes the most attractive case study in research literatures. Nowadays, issues of legal operations concerning service calling, flexible fares, rating system between drivers and users and tax are still in dispute for UBER in Taiwan. It will be very helpful for Taiwan's authorities and taxi operators if users can understand and accept similar services like UBER by using social media mining and sentiment analysis.

Crawler systems were used to collect all possible text data from social media. Opinion mining and sentiment analysis were sequentially performed in this study. A model based on the theory of three-way decisions was used for sentiment analysis which included three sentiment zones: positive, negative and neutral. Results of the empirical study showed that the three topics about UBER operations concerning operation mechanism, regulation and protest and taxation problems were discussed extensively in Taiwan. The hottest topic was mobile platform services offered by UBER which won positive and neutral orientation tendency more than negative one significantly. The whole negative orientation tendency focused on dispute of UBER's illegal operations in Taiwan. It is evident that the first priority for UBER is to apply legal permission and pay tax arrears as soon as possible if UBER attempts to continue operations in Taiwan.

Key Words: *Theory of three-way decisions; Social media mining; Sentiment analysis; UBER*

一、前言

近年來社群媒體 (social media) 發展迅速，「共享經濟」(sharing economy) 的風潮亦隨之興起，媒合個人需求與供給資源之共享商業模式，已成為許多傳統產業之破壞式創新典範。同時，此類新興服務亦產生若干爭議：究竟共享經濟的本質能否為社會帶來更多的資源分享，提供更多賺取額外收入的機會，抑或僅能取代資源供給已飽和的傳統工作，創造另一種新型態的加值酬勞方式？則有待愈來愈多共享經濟的成功模式加以驗證，方可定論。而在當前共享經濟的成功模式之中，UBER 堪稱為國際知名案例。

UBER 於 2009 年在美國成立，其新型態、跨國營運之商業模式，上市之初面臨各國

政府之杯葛與取締，如迴避稅費、保險、駕駛人資格、車輛登記乃至消費者保護等議題，經常發生陳情抗議活動。Cramer 與 Krueger^[1] 在社群經濟的研究指出，在有 UBER 營運的城市，UBER 的空車率較計程車業者低，載客行駛里程亦較計程車高，乘客的待車時間與車資亦皆較計程車業者短且低。此顯示 UBER 的使用效率確實高於傳統的計程車，因此愈來愈吸引許多民眾使用。

2013 年 6 月 UBER 正式進軍臺北，每月乘車次數平均成長幅度迅速，成功因素在於良好的使用者體驗與便捷的 APP 服務，惟隨之而來的是計程車業者的強烈反彈與抗議聲浪。他們認為 UBER 雖有媒合服務的優越性，但實質上已造成勞資雙方權益的不平等，亦衍生出乘客搭乘安全保障、與現有法規不合以及稅費等問題。尤其在社群媒體網路上對 UBER 在臺營運的相關議題，出現大量的發言與回應文本。此一網路輿情事件的發展，可視為觀察我國政府是否準備面對共享經濟環境所產生的新興商業模式，與輔導現有計程車業者轉型的一個里程碑。若能藉由社群媒體挖掘 (social media mining) 而進行情感分析，尋找民眾關注 UBER 合法經營的話題與可能對策，將可作為政府、計程車業者、UBER 與民眾達成共識的參考。

有鑒於此，本研究擬先回顧社群媒體挖掘與情感分析之相關文獻，經由評析而選定情感分析模式；然後使用爬蟲系統蒐集設定期間之 UBER 網路意見文本，轉換成分類資料庫，進行文本意見挖掘與情感分析，並與資訊視覺化工具結合產出圖示，以顯示文本大數據分析之結果；最後，針對民眾關注 UBER 在臺營運所產生的影響議題以及情感認知與觀感進行歸納總結。本研究期望此分析結果，可供政府相關部門制定適時適地的法令與管理機制，促使類似 UBER 創新服務之相關產業可順利在臺灣蓬勃發展，才能發揮「共享經濟」所提倡的共好、利他精神。

二、文獻回顧

2.1 社群媒體大數據與挖掘

隨著網際網路科技時代的來臨，社群網路提供使用者提供想法、評論、分享或藉由社群網路建立社交圈拓展人際關係等行為。民眾透過社群媒體平臺接收與發佈訊息，而產生極大量的大數據 (big data)。大數據又稱為巨量資料，具有資料量大 (volume)、輸入和處理速度快 (velocity)、資料多樣性 (variety) 及真實性 (veracity) 等特性。大數據的計量單位至少是 PB (1000 個 TB)、EB (100 萬個 TB) 或 ZB (10 億個 TB)，其資料型式具有結構化、非結構化和半結構化 3 種類型，將資料轉換為有用的資訊，再將其轉換為具有價值的情報。大數據資料型式可依類型分為下列 3 種：

1. 結構化資料：存放於 RDB (SQL DB) 或資料倉儲 (data warehouse, DW) 中。
2. 非結構化資料：如圖檔、語音檔、影像檔、PDF 檔、Office 檔案、電子郵件、網頁等。

3. 半結構化資料：如 Log 檔、XML 檔、CSV 檔，一般稱為機器資料。

根據統計顯示，全世界結構化數據增長率約為 32%，而非結構化數據增長率則為 63%，預估未來用於智慧產業的大數據，往往是這些非結構化數據。有關社群網路大數據的研究，係著重於社群媒體挖掘 (social media mining) 與情感分析 (sentiment analysis)，前者是在社會互動所產生的社會媒體大數據中，以一種可計算的方式包括數據挖掘、機器學習、社會網路分析、優化演算法、統計學等技術，以表示與衡量社會媒體的虛擬世界，並建立模型以理解其中的交互關係^[2]；後者則藉由社群媒體挖掘，以理解網民欲表達的意見、觀點及喜怒哀樂的情感認知^[3]。社群媒體挖掘與情感分析皆有別於傳統問卷調查方式，近年來在交通、金融、新聞、政府治理、社群互動…等領域之實際案例皆呈爆炸性的成長，茲列舉若干國內外文獻如下：

Drury^[4] 證實社群媒體有別於傳統媒體，除了運用 web2.0 技術並充分發揮線上即時互動功能外，重要的是使用者可藉由網路達到共用影音、照片、圖像、文字、想法、意見、趣事、八卦、新聞等內容，日漸形成社交圈的網路結構。

Manovich^[5] 藉由蒐集社群媒體使用者之圖片、文字、影音、超連結、位置、狀態等資料，進行記錄與數位化之轉換，並應用於各類研究領域而稱為「社群大數據」。他論證由於網路科技的進步發展，使人們開始透過網路，建立與發展社交網路，因此產生社群網路行為。

Chan 等人^[6] 提及社群媒體的類型呈現多樣化，網路使用者最常見之類型為「社群網站」，以 Facebook、Twitter、MySpace、Bebo、LinkedIn 等網站為代表。Wilson 等人^[7] 提出構成「網路社群」的原因，在 Facebook、Twitter 等社群媒體快速發展的時代，民眾習慣透過社群媒體互動交流，民眾之人際關係網絡亦在社群媒體上建立與發展，而具有此共同點之使用者，因此聚集形成網路「社群」(community)。

朱斌好等人^[8] 分析在 2014 年 7 月 8 日到 8 月 11 日間，網路上有關自由經濟示範區的討論文章，並設計與自由經濟示範區有關的「事件」，透過內容觀察法與自動語意技術，觀察網路民意在事件前後的變動情形。結果發現網路大數據分析在輿情觀測上的應用，容易受到其他網路事件之影響，因此建議進行輿情監測時，若考量短時間改變宜採大數據，若考量深度效果則採問卷調查之策略。

蕭乃沂等人^[9] 以臺北市政府的「1999 市民熱線」32,224 筆 1999 市民陳情資料為基礎，測試且驗證大數據分析在改善案件分派效率上之可行性。該研究指出民眾陳情資料為了解民情趨勢的重要管道，政府部門可嘗試導入網路輿情分析，配合現有的民意調查，以提升政策規劃能力。

陶冶中與陳亭愷^[10] 係應用社群運算 (social computing)，建立一包含文本蒐集、文本處理、文本分類、情感分析等步驟之網路輿情分析流程，對於我國實施國道電子計程收費政策後之網路民眾意見，彙整出 ETC 相關議題之輿情趨勢圖，供管理決策者參考。

2.2 情感分析相關研究

情感分析一詞最早源自於 2001 年 Das 與 Chen^[11] 對股票市場留言版之文本訊息，將情感定義為訊息中蘊含的正面 (positive) 與負面 (negative) 意見。2003 年 Dave 等人^[12] 最早使用意見挖掘 (opinion mining) 的概念，認為可針對產品的屬性，挖掘文本中的意見 (好、中、壞)。此後意見挖掘與情感分析大量出現於學術研究，只是情感分析技術主要圍繞新聞報導的長文本，此係因其語法規則完整，易於分析處理。隨著 Web2.0 技術的出現，民眾受到網路的影響遠遠超過任何傳媒方式，且已習慣在社群媒體網路平台發表意見而形成所謂的社群網路流行語、短文本，提高了情感分析的難度。2008 年 Pan 與 Lee^[13] 將情感分析與意見挖掘作為統一術語來處理。2012 年 Liu^[14] 定義情感分析為針對產品、服務、事件、話題等，分析使用者文本中蘊含的意見、情感、態度、情緒等主觀性訊息。

綜合言之，情感分析包括觀點挖掘、評論挖掘、意見挖掘、態度分析、傾向性分析等一系列針對文本表達的人之潛在情感分析過程。然而，文本的情感分析是一個相對較複雜的過程，通常包含三個問題：(1) 文本的主客觀分類，即分出文本內容是主觀的評論或是客觀的陳述；(2) 文本的極性分類 (polarity classification)，又稱為正負向分類，即判斷文本內容是正面的肯定讚賞或是負面的否定批判；(3) 判定文本情感傾向的強弱程度，如：強烈反對、反對、中立、支持、強烈支持，又稱為等級推理 (rating inference)^[15]。本研究對於處理此 3 項問題之相關技術概述如下文說明。

2.2.1 文本情感分析技術

目前文本情感分析技術可概分為以語義規則為基礎、以監督學習為基礎與以話題模式為基礎等三大類，分述如下：

1. 以語義規則為基礎

依照文本情感分析對文本特徵單位的不同，由小至大可分為以下的詞彙、語句及篇章之情感分析：

(1) 詞彙的情感判別

判別詞彙的情感傾向可視為情感分析的基礎，為了定量判別文本中詞彙的情感傾向，通常利用數字區間 $[-1, 1]$ 作為情感權重值，以表示詞彙的褒貶程度。若情感權重值大於 0，則詞彙為褒義詞；情感權重小於 0，則詞彙為貶義詞。中文文本的褒貶傾向判別，主要有基於語料庫和基於語意相似度兩種方法。基於語料庫的方法，主要是利用詞彙之間的連接詞以及統計特性，來判別詞彙的情感傾向。早期學者發現，由於連接詞連接的詞彙之間存在某種情感關聯，例如由連接詞「和」連接的詞彙，情感傾向相同，而由連接詞「但是」連接的詞彙，情感傾向則相反。

有鑒於此，Hatzivassiloglou 與 McKeown^[16] 提出一種判斷情感傾向的方法，假設已知同一句子中某一個詞彙的正負性，透過不同的連接詞關係判斷另一個新詞的正負性，但此方法僅能處理含有連接詞的句法結構，識別模式與處理對象則大受限制。Wiebe^[17] 利用相

似的方法，以相似度分布的詞彙聚類方法，在大語料庫上獲取形容詞性的評價詞彙。然而，前述兩種方法僅將評價詞語的詞性局限於形容詞詞性，卻忽略其他詞性的評價詞語。為避免評價詞語詞性的限制，Riloff 與 Wieber^[18] 經由人工方式，制定出範本並選取種子評價詞語，反覆運算後獲取出詞性的評價詞語。

Turney^[19] 提出點式相互資訊 (point mutual information, PMI) 的方法，判別某個詞彙是否為評價詞語。PMI 方法適用於各種詞性的評價詞語識別，但是較為依賴種子褒/貶詞彙的集合。綜合前述，可知基於語料庫方法最大的優點在於簡單易行，缺點則是可利用的評論語料庫有限，同時評價詞語在大型語料庫中的分布情形並不易歸納。

詞庫相似度方法，則是利用中文辭典 (HowNet) 提供的語意相似度或層次結構，來判別詞彙的情感傾向。Hu 與 Liu^[20] 利用詞之間的同義或反義關係，來推斷新詞的褒貶性，並經由在 WordNet 中建立單詞之間的同義與反義關係，對於未知褒貶性的新詞，只要檢索其在 WordNet 中的同義詞或者反義詞，即可判斷其褒貶性。但某些詞與它們的同義詞不一定具有相同的褒貶性，且通常難以判斷新詞彙的褒貶性強度，故在識別新詞彙的褒貶性僅限於同義關係或反義關係。

(2) 語句的情感傾向判別

語句的情感傾向分析是以詞彙的傾向性分析為基礎，根據上下文關係進行分析，進而得出句子的傾向性。語句的情感分析主要包括對語句的主、客觀性的區分，判別主觀句的褒貶程度。黃心宜^[21] 將文本斷詞後賦予詞性，並將斷詞結果以人工審視後建立主題詞庫、屬性詞庫、否定詞庫及評價詞庫。主題詞庫代表文章主題之特徵詞庫，以名詞居多，屬性詞、否定詞及評價詞庫則是包含各種形容詞、副詞及動詞，用來修飾名詞。斷詞文本結果經語料庫比對後，會根據比對後擷取之意見單元，進行其正負向分數的計算，以表示情感分數。若以「這隻手機很不錯」為例，手機是屬性詞，很是副詞，不錯是形容詞，根據正負向判別給予加權分數，則會以手機 (+1)*不錯 (+1)，即是將屬性詞乘上評價詞的方法，正分即為正向情感，負分即為負向情感。除計算評價分數採用語料庫比對之外，該方法尚包含以語言學為基礎，採用屬性詞擷取後，根據屬性詞對應之詞語進行正負向情感取向加權。

(3) 篇章的情感傾向判別

篇章的情感傾向分析，是指將文本從整體上區分為褒義、貶義或中性取向。楊惠淳^[22] 針對電影評論提出一個情感分析方法，該研究對斷詞文本集合進行特徵量項的選取，以文件頻率 (documentary frequency, DF) 及詞頻-逆向文本頻率法 (term frequency-inverse documentary frequency, TF-IDF) 作為特徵量化以及篩選主、客觀句的方法。實驗採用 5000 筆評論，以中研院 POS 斷詞系統斷詞後，在主客觀分析之前階段時針對主客觀詞進行篩選，以 PMI-IR 方式計算特徵值分數，主觀特徵值為正，客觀為負，反之則視為矛盾予以刪除。實驗部分採用非注音文、火星文之文章，且文章字數大於 50 字，據人工評選後之好評、壞評之文章各 40 篇，以作為實驗對象。實驗結果發現，經主客觀分析後，準確率可達 65%，未經分析則僅 54%，可知主客觀評價分析規則在情感分析上，具有重要的意涵。

另外，使用 PMI-IR 方法較多為英文文章情感判別，中文則較少，而根據詞頻—逆向文本頻率法所計算並建立之情感分析語料庫，在總數 2,500 個特徵詞中，前 2,000 個特徵詞具有辨別的效果，由此可見經詞頻—逆向文本頻率法計算後之特徵量，普遍具有好的權值。

2. 以監督學習為基礎

以監督學習為基礎的情感分析方法，首先以人工標注文本極性，將其作為訓練集，經由機器學習方法建構情感分類器，再進行目標文本的情感分類。Pang 等人^[23]於 2002 年針對電影評論文本，首次將機器學習技術引進文本情感分析。他們用人工標注 752 條負面評論與 1301 條正面評論作為訓練集，分別採用樸素貝式法 (naïve Bayes)、最大熵法 (maximum entropy) 與支持向量機法 (support vector machine, SVM) 對目標文本進行情感分類，結果顯示，機器學習方法可有效提高情感分析的精確度。

3. 以話題模式為基礎

隨著社群網路有關話題 (topic) 的流行，許多學者亦將其應用於情感分析領域，分析使用者對於社會上某個話題或者事件的情感態度。話題模式有 PLSA (probabilistic latent semantic analysis) 模式^[24]與 LDA (latent dirichlet allocation) 模式^[25]，在話題模式的基礎上，增加情感詞變量，同時識別出文檔所談論的話題以及作者的情感傾向。PLSA 與 LDA 模式屬貝氏產生模式。Zhao 等人^[26]於 2010 年提出以最大熵 LDA 為基礎的情感分析方法，其係在 LDA 模式基礎上進行改良，同時識別文本中的評價對象與評價詞語，提高了算法的效率。

Sauper 等人^[27]於 2011 年提出話題與情感結合的聯合模式，挖掘短文片段中的情感傾向性，其採用的話題模式包含隱式馬可夫模型 (hidden Markov model)，亦稱 HMM-LDA 模式。Arjun 與 Liu^[28]於 2012 年提出以半監督學習為基礎之混合模式，容許使用者提供種子詞語而提高情感分析的精確度。

2.2.2 社群媒體情感分析技術

1. 短文本的情感分析

隨著 Twitter、Facebook、新浪微博等社群網路的迅速發展，短小、語法無規則的文本充斥大量噪聲，短文本的情感分析研究即具有重要意義。Go 等人^[29]在 2009 年測試了監督學習算法在 Twitter 短文本的情感分類效果，包含多項式貝式分類、最大熵模式與 SVM。他們採用 Twitter 中的表情符號獲取正面評論與負面評論，因而省去大量的人工標注成本，同時亦提高了訓練集的規模。Pak 與 Paroubek^[30]同樣採用表情符號進行訓練集的蒐集，將其推展至情感分析的三元分類。在情感分類方法上，採用樸素貝氏法分類來獲取初步結果，並採用訊息熵來去除頻繁出現的 n-gram 詞所造成的影響，以達到提高分類結果的目的。Birmingham 與 Smeaton^[31]在 2010 年比較了 SVM 與多項式貝氏算法在長文本與 Twitter 短文本的效果，實驗結果顯示在長文本數據集上，SVM 的情感分類結果優於多項式貝氏法，而在短文本數據集上，多項式貝氏法則較佳。目前使用者利用可社群媒體提供的應用

程式界面 (application programming interface, API) 來抓取大數據，再利用表情符號蒐集數據作為訓練集，然後採用前述監督學習方法來進行情感分類。

2. 以群體智慧為基礎的情感分析

由於社群媒體提供交互功能而增強了使用者情感的交流，因此情感訊息會隨著社群網路持續擴散。Johan 等人^[32] 在 2011 年以 Twitter 大數據為基礎，研究社群媒體中幸福感的同質現象，顯示使用者更傾向於選擇具有相同幸福感指數的人進行交流，其採用無方向加權圖來建構相互關注之關係網路，並以 Jaccard 相似性計算邊的權重。Tan 等人^[33] 針對 Twitter 提出研究使用者層次的情感分析技術，對於每個話題以鏈接結構與內容訊息來建立異構網路圖，然後以半監督學習法中的馬可夫模式，有效提高情感分析的精確度。

2.2.3 情感分析技術之擴展

1. 情感摘要分析

Liu 等人^[34] 於 2005 年首次提出情感摘要技術，針對不同的產品屬性，進行結構化展示。他採用監督關聯規則挖掘方法來挖掘數據集中的頻繁特徵項，經由後續的人工篩選，將隱性特徵映射至顯式特徵，最後基於 WordNet 語義將同義詞特徵進行合併，從而產生最終特徵集合，每個特徵可根據統計結果展示其正面與負面訊息所占比例。

Lu 等人^[35] 針對話題提出了一種基於線上資料庫抽取情感摘要的方法，其假設線上資料庫已包含話題的特徵，如何從大量特徵中選擇有效特徵，以及如何對特徵進行排序而有利於使用者閱讀，即為必須解決的兩個問題。由於此問題為 NP-hard 問題，他們採用貪婪演算法計算而獲得特徵的局部最佳排序。

2. 以移轉學習機制為基礎之情感分析

移轉學習 (transfer learning) 係將數據源分為源領域與目標領域，源領域往往具有大量的標注數據集，而目標領域則無或較少標注樣本。移轉學習的目的，是藉由源領域與目標領域的特徵關聯，將在源領域學習到的特徵直接應用於目標領域中。移轉學習不要求訓練數據與測試數據服從相同的分配，從而能有效在相似領域或者任務之間進行訊息共享與移轉。

Yang 等人^[36] 在 2006 年使用簡單的特徵選擇為基礎的移轉學習策略，選擇在產品評論與電影評價中都具有高排名的詞語作為特徵情感詞，借助電影評價中 2,041 個正面評價與 2,217 個負面評價，成功對產品評價領域進行了情感分類。Blitzer 等人^[37] 在 2007 年針對亞馬遜公司的不同產品分類 (書、DVD、電子產品與廚具) 研究領域的情感分析技術，不僅採用移轉學習方法提高情感分析的精確度，同時研究源領域與目標領域的相關性，在給定目標領域之條件下，如何選擇源領域使其具有最好的移轉學習效果。

綜合上述，目前文本情感分析技術尚處於發展階段，統一成熟的理論體系尚未形成。社群媒體情感分析技術則處於探索期，尤其對於短文本的情感分析方法亟待努力。線上語詞庫愈來愈豐富，亦將在情感分析上帶來更多新的問題與挑戰。一般而言，社群媒體之中

文文本情感分析流程，可概分為爬蟲、辨識、分類與應用等 4 階段，如圖 1 所示。本研究認為在中文文本分類的研究中，多以二元分類，即以正面或負面為判別結果，惟對於文本情感分類中本該存在的猶豫區間，亦即不易劃入正面與負面之文本內容，應可藉由一個中立區間加以分類，來明顯區分不同議題的情感傾向與趨勢，可使情感分類更趨近網路意見的真實性，而保留評論立場的彈性空間，此即在正面、負面之間加入中立情感分類之三元決策理論 (theory of three-way decisions)。

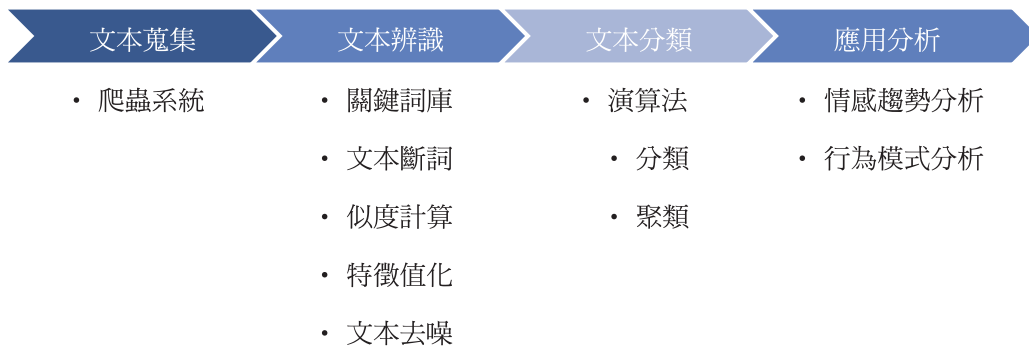


圖 1 社群媒體中文文本之情感分析流程

2.3 三元決策理論

三元決策理論最早是由 Yao^[38] 提出，是傳統二元決策理論的擴展，其思想基礎源自於約略集合 (rough set) 與優勢約略集合理論 (dominance based rough set theory, DRST)，主要目的係將約略集合模型的正向域、負向域及邊界域解釋為 3 種決策結果，分別為接受、拒絕與延遲。經由對粗糙集理論中的正區域、負區域、邊界區域等 3 個區域的語義方面研究，Yao 提出從三元決策角度來解釋粗糙集中的規則提取問題。三元決策規則分別對應物件所屬的正、負、邊界這 3 個區域，根據物件所屬區域不同，分別做出接受該物件屬於目標類、拒絕該物件屬於目標類、對該物件是否屬於目標類不做決策的 3 種決策，並列舉出相對應的決策規則。

三元決策理論可適時模擬人類在解決實際決策問題時之思維模式，並廣泛應用於現實生活的各個領域。在三元決策粗糙集中，依據最小代價做出不同的決策，透過計算物件分類類別的概率與門檻值，將物件劃分到該類別相應的正區域、負區域和邊界域中，分別對應於決策中的接受、拒絕與延遲決策。

由於資訊的不確定性與不完整性，經常無法做到接受與拒絕的二元結果。因此三元決策多了一種延遲的選擇，當資訊不確定，不足以支持接受或拒絕時，應採用第 3 種選擇。三元決策認為延遲選擇亦為一種決策，說明如下：

1. 接受決策：若對象對某標準的滿足度大於或等於某一程度，則採取接受決策，並認為該對象滿足標準。

2. 拒絕決策：若對象對某標準的滿足度小於或等於某一程度，則採取拒絕決策，並認為該對象不滿足標準。
3. 延遲決策：若對象對某標準的滿足度小於某一程度且大於另一程度，則採取延遲決策，並認為尚未能判斷該對象是否滿足標準。

情感分析已成為自然語言處理與人工智慧領域的研究重點之一，而社群媒體挖掘之情感分析是情感認知計算的重要關鍵，因此三元決策理論應納入實際決策思維模式，廣泛應用於現實生活中。目前三元決策理論相關的應用案例可整理如表 1 所示。

表 1 三元決策理論應用領域表

| 應用範疇 | 應用領域 |
|------|---|
| 資訊 | 垃圾郵件過濾 半監督學習 自動分類 E-learning 文本情感分析 |
| 工程 | 產品檢驗 礦場選擇 模型選擇 |
| 管理決策 | 投資組合 政策制定 財務分析 環境管理 論文審稿 |
| 醫療 | 醫療診斷 醫療支援系統 |

資料來源：劉盾^[39]。

在三元決策約略集合中，延遲決策對象係指位於中立域範圍的對象，需要進一步蒐集訊息加以分析，根據新的分析結果與實際情況再做出相應決策，即接受或拒絕。因此三元決策理論可有效處理與分類文本資料，減少錯誤而提高情感分析的準確率。相關研究可彙整如下：

1. 田海龍等人^[40]針對微博文本之觀點句子，進行意見挖掘與輿情分析，他們提出一個基於三元決策的中文微博觀點句子辨識方法，採用支持向量機分類器，以函數計算出微博訊息屬於觀點句子的機率，並將其與傳統支持向量機方法進行比較。結果顯示，基於三元決策的中文微博觀點句子辨識方法獲得較佳的辨識效果。
2. 張志飛等人^[41]認為詞語的情感不確定性來自上下文，句子的情感不確定性來自表達的主題，篇章的情感不確定性來自情緒重疊，因此採用三元決策作為不確定性分析工具，

對詞語、句子、篇章 3 個粒度的文本情感分類，解決了情感分類中的上下文有關、主題與多標記情緒等問題。經由實驗證明，基於三元決策的詞語、句子與篇章情感分類方法，確實可顯著提高情感分類的能力。

3. 王磊等人^[42]提出一個基於主題特徵與三元決策理論相融合的多重標記情感分類方法；首先採用基於主題的情感識別模式，來判斷句子的多元標記情感類別，並在此基礎上結合三元決策理論，以實現對篇章文本的多元標記情感分類。
4. 周哲與商琳^[43]則提出一種新的特徵值提取方法，透過與三元決策理論相結合，並運用在評論意見文本情感分析中，以提高分類器之效率。根據訓練集合建置動態情感詞典，並根據情感詞典提取文本的特徵值，形成特徵量化矩陣。在分類過程中，若分類器對於目標文本的所屬分類確信程度不夠高時，分類器將利用三元決策理論將文本置於中立域中，等待較合適之處理方法。實驗結果在英文電影評論資料集上，基於動態詞典的特徵提取方法可獲得更佳分類準確率，而三元決策理論可將部分樣本放入中立域而提高分類準確率。

回顧上述文獻，可知社群媒體的文本意見、觀點與立場確實錯綜複雜，在情感分析上應保有中立類別，以呈現模稜兩可情況之真實表達，亦即除了肯定或否定以外的不確定區間，應對某些議題的情感態度進行更精確的分析，以達成決策支援輔助的目的。

本研究將針對 UBER 在臺營運後之社群媒體文本來源，進行三元決策的情感分析，首先運用爬蟲系統盡可能蒐集文本資料，將有關 UBER 話題加以分類，以了解各話題評論之集中性與對應的情感趨勢；然後經由人工檢閱文本內容並結合網路情感辭典建置情感詞庫，再將所蒐集 UBER 話題文本的特徵詞，藉由詞頻演算法計算出每篇文本之情感值，運用三元決策理論，將正、負面情感不顯著之意見文本歸納為中立情感，表示較不顯著之情感傾向區間。

三、研究方法

3.1 詞頻演算法

在文本分類方法流程中，為使機器得以辨識文本構成，需將文本依所具有之情感詞特徵項轉換為數值，透過斷詞系統對詞彙出現之數量進行計算，並決定適合表示文本特性之詞彙，以此表現文本所對應之情感傾向程度。詞頻演算法係將篇章文本進行特徵項量化的方法，是一種以統計為基礎且常用於資訊檢索與文本挖掘之加權方法。Luhn^[44]所提出之詞語頻率演算法是透過計算詞彙在文本中出現的頻率，來體現詞彙所具備之重要性，如式(1)所示：

$$tf_{ij} = \frac{n_{ij}}{\sum k n_{kj}} \quad (1)$$

其中， tf_{ij} 表示特徵詞 i 在文本 j 中出現的次數， $\sum kn_{ij}$ 則表示文本 j 中所有出現詞彙的次數總和。詞語頻率演算法在決定特徵量上，具有計算快速且簡單的優點，但缺點則是難以排除低相關高頻率之無用詞彙，以及抓取低出現率但高重要性之詞彙。傳統詞語頻率演算法在面對現今之網路非結構化文本時，易因納入過多不相干詞語，而降低整體特徵項與特徵量的辨識能力。

詞語頻率演算法與逆向文本頻率演算法，在面對文本特徵項量化的計算上仍有受限之處，詞頻－逆向文本頻率演算法，其意涵為若詞彙在某文本出現頻率越高，但在整體文章集合出現頻率較低時，即可表現該詞具有特殊性，故給予較高權重，如式(2)所示：

$$W_{ij} = tf_{ij} \times \log \frac{N}{df} \quad (2)$$

其中， W_{ij} 為詞 T_j 在文本中的權重， tf_{ij} 為詞 T_j 於文本中的次數， d_j 為詞 T_j 於文本集合 N 中，具有詞 T_j 的文本數量。在詞頻－逆向文本頻率計算後，即可完成特徵項量化之步驟，然後可累加所有在意見文本中出現的情感特徵項對應數值，獲得所有相關文本的情感傾向值，再運用三元決策理論加以分類。針對每一篇意見文本，其情感值計算可用式(3)表示：

$$S(s) = \frac{\sum_{f \in F} S(f)}{COUNT(F)} \quad (3)$$

其中， $S(s)$ 表示文本 s 的情感值， F 表示所有母體意見文本所包含的特徵項集合， f 表示單一意見文本的情感特徵項， $S(f)$ 是用詞頻－逆向文本頻率演算法公式計算出的特徵 f 的情感值， $COUNT(F)$ 表示母體文本全部的情感特徵項數量。

3.2 三元決策理論

三元決策規則分別對應物件所屬的正面、負面、中立的 3 個分類區域，根據物件所屬區域不同，分別作出接受該物件屬於目標類（正面）、拒絕該物件屬於目標類（負面）、對該物件是否屬於目標類不做決策的三元決策（中立），並提取出相對應的決策規則。對於決策而言，約略集合模型所需的門檻值參數可由決策損失函數來確定，其運用於社群媒體文本情感分類方法之流程如圖 2 所示。

三元決策理論將文本情感傾向分為 3 類：

1. 正面：文本情感值判斷為顯著正向時歸納為正面文本；
2. 負面：文本情感值判斷為顯著負向時歸納為負面文本；
3. 中立：文本情感值無法明確判斷正、負向時，歸納為中立文本。

3.2.1 基於傳統約略集合理論之三元決策模型

此係由 Pawlak^[45] 提出，主要內容是關於上下近似集合的定義。假設 U 是一個有限的

非空子集合， R 是定義在 U 上的一種等價關係， $apr = (U, R)$ 為近似空間， U 為等價關係 R 下的劃分計為 $U/R = \{[x]_R \mid x \in U\}$ ， $[x]_R$ 是包含 x 的等價類。 $\forall X \subseteq U$ ，其上下近似集合描述如下：

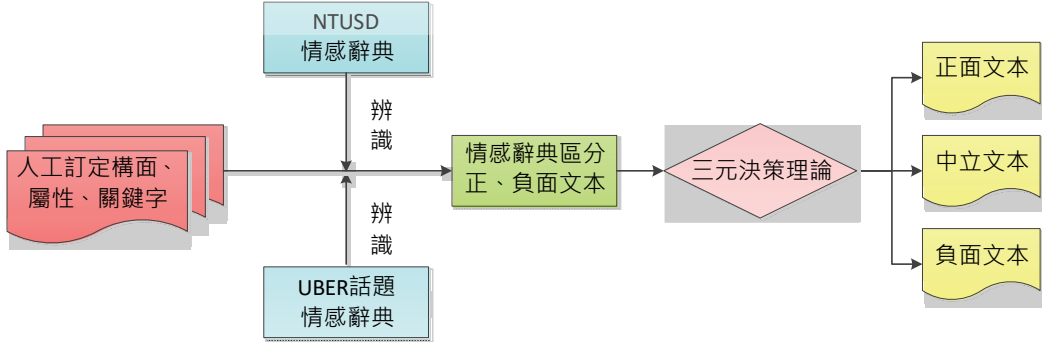


圖 2 結合三元決策理論之社群媒體文本情感傾向分類方法

$$\underline{apr}(X) = \{x \in U \mid [x]_R \subseteq X\} \quad (4)$$

$$\overline{apr}(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\} \quad (5)$$

其中， $apr(X)$ 表示 X 的下近似集合， $\overline{apr}(X)$ 表示 X 的上近似集合，根據條件機率：

$$P(X \mid [x]_R) = \frac{|X \cap [x]_R|}{|[x]_R|} \quad (6)$$

$$P(X^c \mid [x]_R) = 1 - P(X \mid [x]_R) \quad (7)$$

上下近似集合可以表示如下形式：

$$\underline{apr}(X) = \{x \in U \mid P(X \mid [x]_R) \geq 1\} \quad (8)$$

$$\overline{apr}(X) = \{x \in U \mid P(X \mid [x]_R) > 0\} \quad (9)$$

上下近似集合把論域 U 分成 3 個部分，正域 $POS(X)$ ，邊界域 $BND(X)$ 和負域 $NEG(X)$ 。根據上下近似集合的概念，可得到基於傳統約略集合理論的三元決策模型的三元決策規則，分別表示如下：

$$POS(X) = \underline{apr}(X) \quad (10)$$

$$BND(X) = \overline{apr}(X) - \underline{apr}(X) \quad (11)$$

$$NEG(X) = U - \overline{apr}(X) \quad (12)$$

亦可以下列形式表示：

$$POS(X) = \{x \in U \mid P_r(X|[x]_R) \geq 1\} \quad (13)$$

$$BND(X) = \{x \in U \mid 0 < P_r(X|[x]) < 1\} \quad (14)$$

$$NEG(X) = \{x \in U \mid P_r(X|[x]) \leq 0\} \quad (15)$$

對於 $\forall x \in U$ ，當 $x \in POS(X)$ 時， x 一定屬於 X ；當 $x \in BND(X)$ 時， x 還不能確定是否屬於 X ，需要延遲決策；當 $x \in NEG(X)$ 時， x 一定不屬於 X 。對於一個樣本 x ，可以將其劃分到 3 個區域，亦即其有 3 個決策規則。此顯現出三元決策的基礎概念。

3.2.2 基於區間值模糊集合之三元決策模型

設 $A = [A^-, A^+]$ 是 U 的一個區間值模糊集合，取 $E(A)(x) = A^{(m)}(x)$ ，則 $(U, Map(U, I), [0, 1], E)$ 是一個三元決策空間。若 $0 \leq \beta < \alpha \leq 1$ ，則三元決策如下：

$$\text{正面域：} ACP_{(\alpha, \beta)}(E, A) = \{x \in U \mid A^{(m)}(x) \geq \alpha\}$$

$$\text{負面域：} REJ_{(\alpha, \beta)}(E, A) = \{x \in U \mid A^{(m)}(x) \leq \beta\}$$

$$\text{中立域：} UNC_{(\alpha, \beta)}(E, A) = \{x \in U \mid \beta < A^{(m)}(x) < \alpha\}$$

設 $A = [A^-, A^+]$ 是 U 的一個區間值模糊集合，取 $E(A)(x) = A(x)$ ，則 $(U, Map(U, I), I, E)$ 是一個三元決策空間。若 $0 \leq \beta < \alpha \leq 1$ ， $\alpha = [\alpha^-, \alpha^+]$ ， $\beta = [\beta^-, \beta^+]$ ，則區間值三元決策如下：

$$\text{正面域：} ACP_{(\alpha, \beta)}(E, A) = \{x \in U \mid EA(x) \geq \alpha\}$$

$$\text{負面域：} REJ_{(\alpha, \beta)}(E, A) = \{x \in U \mid EA(x) \leq \beta\}$$

$$\text{中立域：} UNC_{(\alpha, \beta)}(E, A) = \{x \in U \mid E(A)(x) \geq \alpha\} \cup \{x \in U \mid E(A)(x) \leq \beta\}$$

3.3 求解模糊門檻值

解模糊的方法有很多，其中最廣為使用的有：區域中心法 (center-of-area, COA) 來得到解模糊 (defuzzifier value) 或最佳非模糊績效值 (best non defuzzifier performance, BNP)。例如，若以重心法 (center-of-gravity, COG) 求取 BNP，則對模糊數 (L, M, U) 而言，其 BNP 定義如下：

$$BNP = DF = \frac{(U_i - L_i) + (M_i - L_i)}{3} + L_i \approx \frac{U_i + M_i + L_i}{3}, \forall i \quad (16)$$

由於 HAHP 或 HDHP 均以語意變數來評分，不像一般的明確值能夠直接計算，因此要利用解模糊化的方式將模糊數和語意變數轉變為明確值。目前，解模糊數的方法有很多，常見的解模糊化的方法包括：重心法、最大平均法 (mean of maximum method)、 α -截集法 (α -cut method)、中心平均解模糊化法 (center average defuzzy) 與模糊積分之排序。

3.3.1 重心法

重心法即是找出三角形面積中心點之概念，將模糊數的面積中心點視為其代表值。若論域 U 為實數域中的有界集合，則 U 中的模糊集 $A: \mu_A(x)$ 的模糊數重心為：

$$DF = \frac{\int_U u_A(x) \times x dx}{\int_U u_A(x) dx}, \text{ 其中 } \int_U u_A(x) dx \neq 0 \quad (17)$$

若 $U = [a, b]$ 時，則模糊數重心為：

$$DF = \frac{\int_a^b u_A(x) \times x dx}{\int_a^b u_A(x) dx}, \text{ 其中 } \int_a^b u_A(x) dx \neq 0$$

若 $U = \{x_1, x_2, \dots, x_n\} \subset R$ 時，模糊數重心為：

$$DF = \frac{\sum_{i=1}^n u_A(x_i) \times x_i}{\sum_{i=1}^n u_A(x_i)}, \text{ 其中 } \sum_{i=1}^n u_A(x_i) \neq 0 \quad (18)$$

當模糊數為三角模糊數時，則重心法 (BNP) 公式 $DF = \frac{\int_U u_A(x) \times x dx}{\int_U u_A(x) dx}$ ，可轉換成下列

線性式 (19)，重心法示意圖如圖 3 所示：

$$DF = \frac{(M_i - L_i) + (U_i - L_i)}{3} + L_i \cong \frac{[L_i + M_i + U_i]}{3} + \frac{3 \times L_i}{3} \cong \frac{L_i + M_i + U_i}{3}, \forall i \quad (19)$$

其中 DF_i ：解模糊化後的明確值；

U_i ：三角模糊數的最大值；

M_i ：三角模糊數的中間值；

L_i ：三角模糊數的最小值。

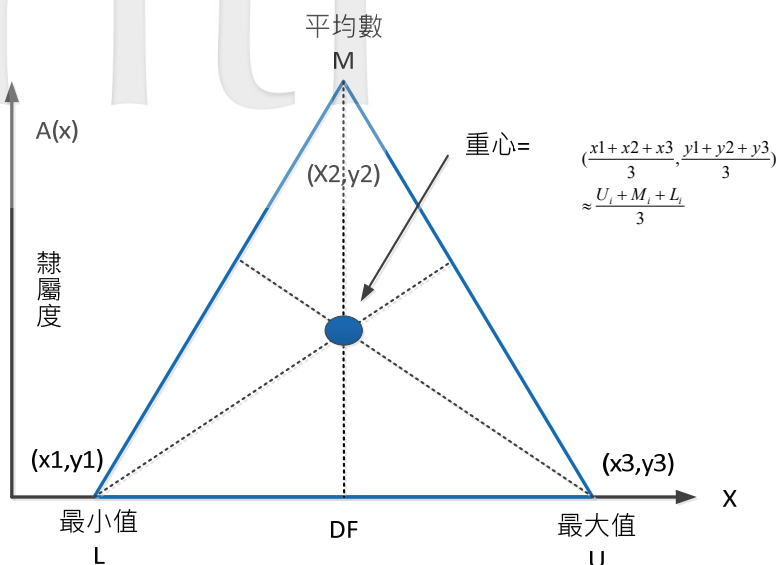


圖 3 三角形模糊數重心示意圖

四、實證分析

本研究運用網路爬蟲蒐集論壇、新聞網站、社群網站與部落格中有關 UBER 之文本資料，利用文本挖掘技術，計算文本資料中出現之情感特徵項次數，作為判斷情感傾向之依據，輔以討論量觀察不同時間區段的輿情尖離峰，再利用三元決策理論計算中立情感區間值，藉此將情感強度不足以被劃分為正、負情感傾向之文本歸納至其類別，以加強各話題發生事件時之情感判別準確性，更加符合實際社群網路評論行為的分類模式。

因考量 UBER 相關話題與追溯過往之相關文本，以了解民眾在網路上的發文內容，以及 UBER 營運話題是否引起廣大民眾熱烈討論，故本研究將聚焦於討論量爆量之相關話題事件，以直觀 UBER 營運話題之情感趨勢為主，而排除低度討論量之文本。

4.1 資料處理流程

本研究之實證過程可分下列 3 階段：

1. 文本資料預處理

將 UBER 營運話題相關之文本資料，藉由爬蟲程式從各社群媒體討論區取得原始母體樣本，經過人工檢視每筆文本資料內容之後，再進行篩選，然後刪除關聯度低的無效樣本，並建立專屬 UBER 的營運議題與其對應的情感特徵值詞庫。藉由自行開發專屬的詞頻—逆向文本頻率計算程式，計算出所有在母體樣本中出現過的情感特徵值所代表的數值，並逐一取代每篇文本出現過的特徵項為數值並加以累加計算，計算出每一篇 UBER 相關的文本

的情感傾向值，作為後續分類之用。

2. 三元決策之情感傾向分析模型

以往研究常利用二元分類方法，將情感詞庫分類文本之情感傾向劃分為正面、負面兩類。本研究係以三元決策理論為基礎，將文本之情感傾向劃分為 3 個類別，運用三元決策理論之模糊區間與解模糊化之重心法，計算求得中立門檻區間的分界門檻值，將情感值區間劃分為正面、中立與負面 3 類，再將所有文本之情感值依數值由大至小排列，利用重心法計算之模糊區間門檻值區間，來劃分中立之情感區間。

3. 情感趨勢分析之視覺化圖式

本研究以視覺化圖式方式呈現 UBER 相關營運話題的情感趨勢，並說明社群媒體文本在不同時間的尖離峰情感趨勢及對應現實輿情事件的狀態。藉由分析不同話題的討論聲量與不同時間區段的情感傾向值，可得知目前 UBER 在臺灣營運話題之民眾觀感與在意的話題，再藉由三元決策的門檻值，清楚劃分正面、中立以及負面傾向的營運話題，分析其情感趨勢，提供相關單位作為決策輔助之參考。

4.2 整體趨勢分析

本研究運用爬蟲程式蒐集有關 UBER 之社群媒體文本，時間範圍自 2015 年 7 月 1 日至 2016 年 9 月 13 日，文本資料範圍包含社群網站、討論區、部落格、新聞 Facebook 留言、BBS 等主要核心社群媒體。經由自動摘取討論及回應文章，再逐篇過濾評論內文，挑選出具有分析價值的有效文本。基於資料保護之隱私原則，本研究並未針對此兩類文章中表明是否曾搭乘 UBER 的貼文者或回文者進行偵測與追蹤，亦無從推估貼文者或回文者是否為 UBER 公司所為。再者，由於交通部對於 UBER 迄今仍採非法取締措施，多數貼文與回文者並不主動表明自己是 UBER 的實際體驗者，而多表達對於 UBER 的觀點、意見與看法。初步彙整民眾對 UBER 在臺營運話題的討論與回應數量，如表 2 所示。總討論文章數量共計 798 篇，回文數量則達 39,474 篇，可知相關營運話題的文章數量越多，回應文本數量則越多。

前三大討論聲量的營運話題分別為營運制度、取締與抗議以及稅務與收費，文章數量分別為 409、189 與 53 篇，回文數量分別為 22,497、10,508 與 1,860 篇，其中又以營運制度的討論文章與回文數量最高，次多者則是取締與抗議的話題，有超過 2 倍以上的討論數量。

至於討論聲量較低的話題則有大數據應用、計程車司機個人素質及商業模式，文章數量分別為 16、8 和 3 篇，回文數量分別為 161、135 和 43 篇。此顯示民眾對於主要營運爭議話題，如營運制度、合法性和稅務問題討論度較為熱絡，至於加值應用服務內容以及計程車駕駛人素質為主的討論串，討論量則較少。

本研究在進行實際文本蒐集與分類調查時，審閱過全數母體有效樣本後，初步將 UBER 在臺營運話題區分為四大類別，分別為：計程車產業、共乘平台之營運機制與應用技術、

表 2 UBER 在臺營運話題之文本討論數量與回應數量

| 議題類別 | 文章數量 | 回文數量 |
|------------|------|--------|
| 營運制度 | 409 | 22,491 |
| 取締與抗議 | 189 | 10,508 |
| 稅務與收費 | 53 | 1,860 |
| 乘客安全保障 | 52 | 1,567 |
| 計程車硬體與個人行為 | 45 | 1,438 |
| 創新應用 | 23 | 1,271 |
| 大數據應用 | 16 | 161 |
| 計程車司機素質 | 8 | 135 |
| 商業模式 | 3 | 43 |
| 總數 | 798 | 39,474 |

資料來源：本研究彙整

UBER 營運爭議、共享經濟發展方向。在此四大類別主題下，又再細分為 9 個主要核心討論話題，每篇有效的文本話題分類係依據對應的自建詞庫統計數量。自建詞庫是藉由統計軟體計算頻次彙整而得，單篇文本（含回文內容）議題歸納則是經由關鍵字詞頻統計，以特定議題的關鍵字出現頻次最高者所定義。本研究定義 UBER 在臺營運之構面、話題與對應之關鍵詞庫如表 3 所示。

表 3 UBER 在臺營運之構面、話題與對應之關鍵詞庫

| 構面類別 | 話題 | 自建詞庫 (列舉) |
|----------------|--------------|--|
| 計程車產業 | 計程車司機素質與駕駛行為 | 乘客體驗、跳表、繞路、車型、煙味、廣播、政治立場、司機聊天、拒載、濫收車資、服裝儀容不佳、喊價、繞路、投訴、客訴、司機素質、車輛外觀 |
| 共乘平台之營運機制與應用技術 | 營運制度 | 經營優勢、糾紛處理、P2P、制定專法、評鑑、社群、App、平台、叫車服務、導航、車輛追蹤、GPS 整合、線上付費、評價機制、法律規範、資訊服務、法令、里程 |
| | 大數據應用 | |
| UBER 營運爭議 | 稅務與收費問題 | 繳稅、稅收、跨境收取車資、規避稅捐、違法經營、抗爭、遊行、罷工、計程錶、費率、特許行業、執照、乘客權益、開罰、罰金、剝削、罰款、消費糾紛、第三方保險、違規派遣、稅租、無牌、白牌 |
| | 取締與抗議 | |
| | 乘客安全保障 | |
| 共享經濟發展方向 | 商業模式 | 閒置資源、媒合服務、產業衝擊、商業模式、共享、創新科技、共乘、分享經濟、預約、協作、資源共享、偏遠地區 |
| | 創新應用 | |

資料來源：本研究彙整

4.2.1 社群媒體文本整體情感分析

本研究以人工檢閱，過濾刪除明顯無關之討論文章，保留資料時間 2015 年 7 月 1 日至 2016 年 9 月 13 日止之 UBER 相關文章，然後利用文本挖掘技術分析過濾後的文本內容。針對討論與回應文本之重要關鍵字、話題相關關鍵字與情感關鍵字等，分別計算其討論總量與情感傾向總量，並以週為單位加總每日討論量與情感傾向值，藉此綜觀民眾討論熱度高峰時之事件發生原因、事件內容、討論內容與觀點、輿情狀態是否持續、事件消散原因等，釐清民眾對 UBER 營運話題事件之支持程度，如圖 3 所示。

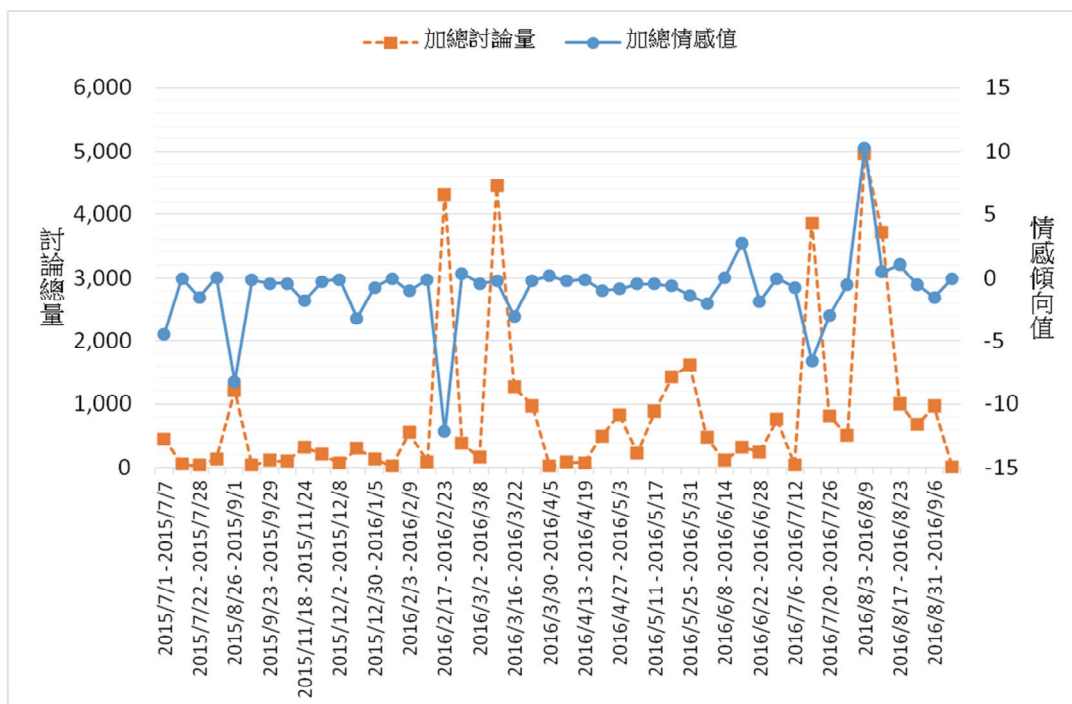


圖 3 UBER 在臺營運話題之討論總量與情感總量趨勢圖

以整體趨勢而言，UBER 話題在 2015 年討論量較少，至 2016 年為止有明顯之峰值事件，分別為 2016 年 2 月 17 日至 23 日、3 月 9 日至 3 月 15 日、7 月 13 日至 7 月 19 日以及 8 月 3 日至 8 月 9 日的 4 個峰值。由情感趨勢圖可知情感傾向累加值多半落在 $[-1.5, 1]$ 之間，由此可知民眾針對 UBER 在臺營運話題之情感傾向多為中立偏向正向。若輔以人工檢視其內容，可得知民眾對於 UBER 所提供之服務品質有一定程度的支持，再檢視其負面情感傾向文本，發現民眾多對 UBER 營運合法性與正當性爭論較大，茲彙整成以下之表 4 至表 7：

- (1) 政府開始正式取締開罰 (2016 年 2 月 17 日至 2 月 23 日)

表 4 2016 年 2 月 17 日至 2 月 23 日尖峰聲量事件分析

| 輿情趨勢 | 開始增長 | 達到尖峰 | 漸趨下降 | 消散 |
|--------|--|----------|----------|----------|
| 發生時間 | 2 月 17 日 | 2 月 22 日 | 2 月 23 日 | 2 月 23 日 |
| 情感極值 | 正向情感值 | | 負向情感值 | |
| | 最小值 | 最大值 | 最小值 | 最大值 |
| | -0.0233 | 0.1518 | -0.5242 | -0.3477 |
| 尖峰時期 | 2 月 18 日~2 月 22 日 | | 平均情感 | -0.11462 |
| 爭議事件 | 1.新聞報導法院認定 UBER 違法 2.新聞報導交通部將嚴格吊銷 UBER 執業司機 3.民眾以自用車執業 UBER 遭吊照與開罰 4.高雄開罰 UBER 司機並吊照 | | | |
| 事件詳細內容 | 媒體於 2 月 17 日報導法院判決 UBER 司機違法執業，交通部勝訴，媒體並於同日報導交通部將修法嚴格吊銷 UBER 執業司機車牌，2 月 18 日網路論壇中開始出現零星討論，2 月 19 日媒體報導 UBER 來臺 3 年，法院認證違法新聞後，自 20 日起至 22 日止，討論量進入尖峰區間，民眾於網路論壇中熱烈討論 UBER 在臺營業之合法性，及比較 UBER 與我國計程車搭乘之服務體驗，民眾對於 UBER 提供之服務、車資、叫車便利性、車況與硬體設備等有相當正向的支持，惟難認同 UBER 在臺違法營業的事實現況。 | | | |
| 輿情消散事件 | UBER 相關討論持續約達一周，民眾關注 UBER 在臺營業的合法性，無論反對或支持者普遍認為 UBER 雖服務優良，但合法營業應為最優先事項，2 月 22 日、23 日媒體報導高雄開出首宗 UBER 違法營業罰單後，顯示政府機關已完全定調 UBER 在臺屬營業違法，故相關討論量呈現消散狀態。 | | | |

從表 4 可知，民眾針對 UBER 的討論可分為兩個部分：服務體驗與營業合法性。經檢閱該時期相關討論文本，其內容多著重於 UBER 之服務體驗，諸如司機素質、車況與硬體設備、叫車便利性、後續服務等項目皆備受使用者肯定，相較於現有計程車業者，司機素質不一、車況普遍不佳、煙味重、繞路、拒載、無申訴管道、電話叫車久候、車源不足等情況，皆為民眾經由比較而肯定 UBER 服務的原因，即使知曉 UBER 尚無合法營業執照，民眾表現出的平均情感傾向亦為中立偏向正面，驗證了多數民眾對於 UBER 的服務抱持正面態度。

(2) 社群媒體警示 UBER 營運非正當 (2016 年 3 月 9 日至 3 月 15 日)

對照表 4 與表 5 相關討論之尖峰時期，可發現 UBER 在臺營運之服務體驗雖備受肯定，但營業合法性仍是難以忽視的首要議題。比較 2 月 17 日至 23 日與 3 月 9 日至 15 日兩個爆量尖峰時期，輿情消散皆可對應至媒體報導加強取締 UBER 的新聞，顯示民眾在政府定調 UBER 違法營業之後，多數民眾已認同 UBER 在臺違法營運的事實，因此討論熱度降低。

表 5 2016 年 3 月 9 日至 3 月 15 日尖峰聲量事件分析

| 輿情趨勢 | 開始增長 | 達到尖峰 | 漸趨下降 | 消散 |
|--------|---|---------|----------|----------|
| 發生時間 | 3 月 9 日 | 3 月 9 日 | 3 月 11 日 | 3 月 15 日 |
| 情感極值 | 正向情感值 | | 負向情感值 | |
| | 最小值 | 最大值 | 最小值 | 最大值 |
| | -0.0222 | 0.3827 | N/A | N/A |
| 尖峰時期 | 3 月 9 日~3 月 11 日 | | 平均情感 | -0.01341 |
| 爭議事件 | 1. 民眾在網路論壇中轉貼舊聞報導，警示其他 UBER 執業司機目前不具合法性，提醒政府單位取締 2. 民眾轉貼舊聞討論 UBER 合法與否 | | | |
| 事件詳細內容 | 延續 2 月中之討論趨勢，民眾在舊聞報導中討論 UBER 目前在臺營業之合法問題，且多認為合法營業是最優先之事項；此外，民眾討論合法性議題時，亦常列舉如 UBER 受法規管制後，在車資計算、稅費、乘客保險方面將可更受到保障，顯見討論 UBER 在臺執業問題觀點全面；整體觀之，事件雖有較多討論聲量，但在情感傾向呈現明顯更為正向的狀況。 | | | |
| 輿情消散事件 | UBER 營業合法性雖仍備受民眾爭議討論，但從平均情感上來看，民眾普遍無強烈的顯著情感波動；媒體於 3 月 14 日報導政府單位將加重取締 UBER，顯示政府單位定調 UBER 違法營業且更加重視，故使民眾爭議討論消散。 | | | |

(3) UBER 抗告勝訴引發計程車公會包圍立法院(2016 年 7 月 13 日至 7 月 19 日)

表 6 2016 年 7 月 13 日至 7 月 19 日尖峰聲量事件分析

| 輿情趨勢 | 開始增長 | 達到尖峰 | 漸趨下降 | 消散 |
|------|---|----------|----------|----------|
| 發生時間 | 7 月 13 日 | 7 月 13 日 | 7 月 14 日 | 7 月 18 日 |
| 情感極值 | 正向情感值 | | 負向情感值 | |
| | 最小值 | 最大值 | 最小值 | 最大值 |
| | -0.0198 | 0.4213 | -1.1780 | -0.3217 |
| 尖峰時期 | 7 月 13 日 | | 平均情感 | -0.0641 |
| 爭議事件 | 1. 媒體報導 7 月起六都加強取締 UBER 2. 交通部長賀陳旦表示將於 2 個月內推出臺灣版 UBER 3. 媒體報導經濟部擬開放 UBER 合法上路以刺激消費 4. 媒體報導 UBER 遭交通部勒令停業抗告勝訴 5. 計程車抗議 UBER 違法營業包圍立法院 6. 媒體報導投審會將於 1 個月內撤銷 UBER 執照 | | | |

表 6 2016 年 7 月 13 日至 7 月 19 日尖峰聲量事件分析 (續)

| | |
|--------|--|
| 事件詳細內容 | 7 月 13 日起，媒體大幅度報導 UBER 相關新聞，使自 3 月起沉寂之 UBER 議題再受民眾關注。主要事件包含 UBER 抗告交通部勒令停業勝訴、交通部長賀陳旦表示將推出臺灣版 UBER、交通部支持創新應用、經濟部擬開放 UBER 合法上路等，對於 3 月起被政府機關定調為違法營業之 UBER 在其合法性與營運上又產生新的可能性，合法性議題再度被民眾開始討論。相較過去討論，此時間段民眾針對 UBER 合法性與否的討論增加了「是否該對 UBER 進行專法」或是「修正 UBER 營運方式並歸入現行法律列管」以許可其營業、創新共享經濟應用議題，以及不公平競爭與計程車司機反 UBER 三大議題，在合法性討論上，由於媒體報導法院判決 UBER 遭勒令停業勝訴，使民眾對於 UBER 在臺合法營業許可上看見轉機，促使民眾重新討論是否有其他方式可允許 UBER 在臺營業，而交通部長亦肯定共享經濟是未來趨勢，除將協助計程車建立網路平台外，亦希望 UBER 能夠合法化，使民眾對於以 UBER 合法營業為前提之支持討論更為明顯。 |
| 輿情消散事件 | 7 月 13 日前後是自 3 月後，首度出現的多事件發生時間點，無論是法院判決 UBER 再停業抗告勝訴、計程車抗議、交通部、經濟部與投審會採取的各種措施上，雖給民眾新的討論議題，但其根本之合法營業許可問題仍未解決，故當日雖有強烈的討論，但就如同 2 月與 3 月之事件消散議題相同，民眾仍有「合法營業為 UBER 優先事項」之共識，故在新議題持續出現情況下聲量自然消散。 |

(4) UBER 抗告勝訴引發計程車公會包圍立法院 (2016 年 8 月 3 日至 8 月 9 日)

表 7 2016 年 8 月 3 日至 8 月 9 日尖峰聲量事件分析

| 輿情趨勢 | 開始增長 | 達到尖峰 | 漸趨下降 | 消散 |
|--------|---|---------|---------|---------|
| 發生時間 | 8 月 3 日 | 8 月 5 日 | 8 月 7 日 | 8 月 9 日 |
| 情感極值 | 正向情感值 | | 負向情感值 | |
| | 最小值 | 最大值 | 最小值 | 最大值 |
| | -0.0198 | 0.8112 | -0.3288 | -0.3114 |
| 尖峰時間區段 | 8 月 5 日 | | 平均情感 | 0.1185 |
| 爭議事件 | 1.媒體報導 8 月 3 日投審會要求 UBER 撤資 2.媒體報導 8 月 3 日晚間政院暫緩 UBER 撤資，決議在交通部計程車措施 3.公布前暫緩撤資案 4.媒體報導交通部將繼續對 UBER 加強取締 | | | |
| 事件詳細內容 | 投審會要求 UBER 撤資新聞發出後，在新聞網站中之網友回文區受到民眾熱烈討論，民眾普遍認為 UBER 可留在臺灣，但應有合法營業許可，可透過 UBER 與計程車合併、專法管理方式減輕其合法營運之困難性，政院擋下 UBER 撤資新聞除考量避免外國企業誤會我國不歡迎外資投資外，將透過輔導 UBER 進到計程車行業義務併行建置計程車叫車平台，以提升計程 | | | |

表 7 2016 年 8 月 3 日至 8 月 9 日尖峰聲量事件分析 (續)

| | |
|--------|---|
| 事件詳細內容 | 車產業競爭力，民眾對此表示認同，亦表示相較國外已有多數 UBER 營業成功案例，我國採取撤資 UBER 方式是否妥當仍有討論空間。整體而言，本次尖峰聲量事件仍未對 UBER 在其合法性議題與民眾支持傾向帶來改變，合法性仍是民眾關注的重點，值得注意的是本次事件是第一次出現完全正向的平均正面情感，顯見民眾對 UBER 的支持態度已相當明確。 |
| 輿情消散事件 | 暫緩撤資並協助計程車建立平台，輔導 UBER 合法營業新聞出現，政府機關措施與民眾認同達到共識，輿情消散。 |

若合併表 6 與表 7 兩個時間點，可知民眾在 3 月後對 UBER 的支持漸趨明顯，政府對 UBER 營業合法性的堅持與配套措施方向亦逐漸明朗。在 7 月到 8 月之事件討論中，可發現相較過去討論內容僅以合法性與否為觀點，專法管理、修正營運方式併入計程車管理、共享經濟趨勢以及外資投資等觀點開始受到民眾重視，顯示民眾對於新興服務行業在臺灣的接受度相當高。4 個爆量事件時間點之整體平均情感值亦呈現明顯的開低走高，可發現民眾隨著時間的推移、資訊接收程度以及政府態度與措施，對於 UBER 若能在合法營業之前提下，有一定程度的正面支持態度。

4.2.1 UBER 在臺營運三大核心議題

本研究為獲得 UBER 在臺營運話題文本之三元類別結果，須計算出中立門檻區間，方可進行文本分類。本研究係以三元決策理論為基礎，將正面區間門檻值之條件分為：正面情感傾向之最大值 (P_{max})、中間值 (P_{avg}) 和最小值 (P_{min})；負面區間門檻值之條件亦然為，負面情感傾向之最大值 (N_{max})、中間值 (N_{avg}) 和最小值 (N_{min})，如表 8 所示。

中立情感係透過正、負面情感詞庫為特徵值為計算之依據，並以詞頻演算法計算社群媒體文本的情感值，當情感值透過被判斷為不明確之正面情感或負面情感時，則將其視為中立情感傾向。

表 8 UBER 在臺營運話題之情感傾向區間門檻值

| 正面情感傾向門檻值 | 負面情感傾向門檻值 |
|---------------------|---------------------|
| $P_{max} : 0.0406$ | $N_{max} : -0.0878$ |
| $P_{avg} : 0.0077$ | $N_{avg} : -0.2387$ |
| $P_{min} : -0.0225$ | $N_{min} : -0.6508$ |

本研究在計算三元決策情感分析分類模型之通用門檻值時，由於中立文本在 $P_{max} - N_{min}$ 區間之分類準確率最高，而 $P_{min} - N_{min}$ 區間為正面為本分類準確率最高之情感區間。至於負面文本之分類準確率在不同分類狀態之準確度無顯著差異，因此本研究以 $P_{max} - N_{min}$

與 $P_{min}-N_{min}$ 門檻區間，利用重心法進行模糊門檻值的求解計算過程，最後計算出 UBER 文本之三元分類通用門檻值，其中利情感傾向門檻區間介於-0.0225 與-0.3196 之間，文本之情感值大於-0.0225 即視為正面情感傾向文本，文本情感值小於-0.3196 即為負面情感傾向文本。計算出的三元決策分類 (正面、中立、負面) 門檻值如表 9 所示，依照此數值將所蒐集知社群媒體文本區分為三大情感傾向類別，詳細分類之文本數量如表 10 所示。本研究以文本數量前三大話題，分別說明 UBER 在臺營運話題之情感趨勢。

表 9 UBER 在臺營運話題之情感傾向區間門檻值

| | |
|------------|----------------------|
| 正面情感之門檻值 | -0.0225 |
| 中立情感之門檻值區間 | -0.0225 與 -0.3196 之間 |
| 負面情感之門檻值 | -0.3196 |

表 10 UBER 在臺營運話題之文本情感傾向三元分類數量表

| 話題類別 | 正面篇數 | 中立篇數 | 負面篇數 |
|------------|------|------|------|
| 營運制度 | 219 | 155 | 35 |
| 取締與抗議 | 68 | 91 | 30 |
| 稅務與收費 | 19 | 10 | 26 |
| 乘客安全保障 | 22 | 24 | 5 |
| 計程車硬體與個人行為 | 14 | 15 | 15 |
| 創新應用 | 16 | 4 | 3 |
| 大數據應用 | 9 | 5 | 2 |
| 計程車司機素質 | 5 | 2 | 1 |
| 商業模式 | 2 | 1 | 0 |
| 總數 | 374 | 307 | 117 |

資料來源：本研究彙整

1. 營運制度

營運制度是 UBER 在臺營運話題中討論文章最多的分類，達 409 篇，其中正面文本 219 篇、中立文本 155 篇和負面文本 35 篇，整體情感傾向偏向正面為主。正面與中立文本討論集中的項目依數量多寡為：動態調整費率、司機評價機制、圖資定位資訊應用、優質使用介面與網路媒合、清楚的導航路線、回應顧客需求迅速等，此亦顯示出臺灣既有計程車業者在服務新穎之技術方面不及 UBER。

負面文本多針對 UBER 在臺營運是否合法的話題討論最多，如法令、法規制度、制定專法、牌照數量、商業登記等關鍵字項目，可知 UBER 在臺營運之合法性最受爭議。從 2016 年 3 月 5 日至 11 日之尖峰討論時期爆發出 UBER 的合法性討論的數量，亦為政府關

注焦點。民眾在此期間，已透過新聞與社群媒體討論，呈現出 UBER 在臺營運並非合法之普遍意見，且歷經 2 個月的討論延燒，此事件雖有單日大量討論，但後續討論量較小，並隨著媒體報導熱度的下降而消散。圖 4 為 UBER 在臺營運制度話題之三元決策情感趨勢圖。

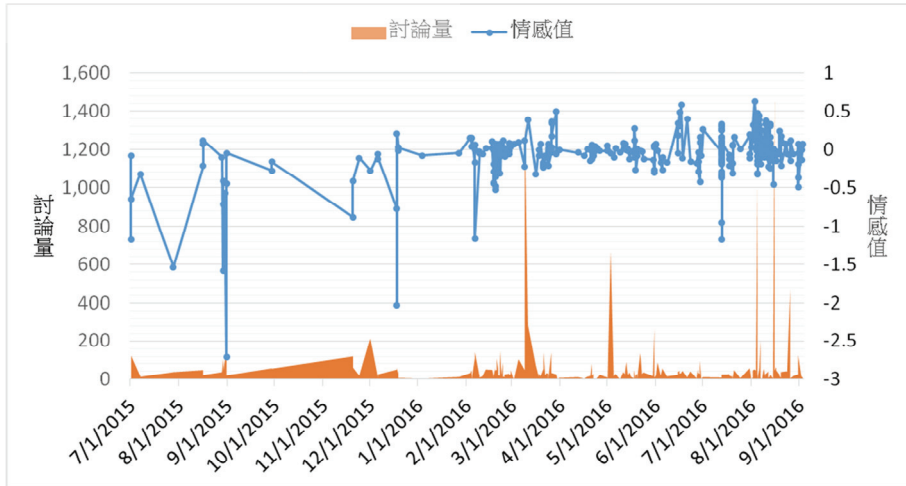


圖 4 UBER 在臺營運制度話題之三元決策情感趨勢圖

2. 取締與抗議

取締與抗議議題是 UBER 在臺營業討論中討論文章次多的分類，達 189 篇，其中正面文本 68 篇、中立文本 91 篇和負面文本 30 篇，整體情感傾向偏向中立。中立文本討論集中的項目包括：勒令停業、罰款、罰金、不公平競爭等；而負面文本討論多針對 UBER 在臺營運尚未具備與計程車產業繳稅的相同基礎，亦多次提及未有核准執照、白牌車違規派遣、營運漏洞、非法駕駛、剝削勞工等關鍵字。

在此議題分類下僅有一次爆量事件，時間為 2016 年 3 月 9 日當日，其餘討論量則平均分布在各時間點中，且皆未引起太大量的討論。該事件為民眾在論壇中討論目前 UBER 的優良服務品質及政府取締執法的手段是否得宜，從討論中衍生部分民眾對於我國現行計程車服務與司機素養多有不滿，諸如：車資收費較貴、拒載、繞路、司機個人修養、車況品質與難以叫車等。

從爆量事件當日之情感趨勢來看，亦可發現民眾對 UBER 在臺違法營業遭取締與抗議，並未表現出太過負面的情緒，比較總體趨勢分析可知民眾雖支持 UBER 提供的優良服務，但因違法而產生的取締問題仍應解決，且未表示出強烈的不滿，顯見民眾普遍認同 UBER 合法營運應為最優先議題，圖 5 為 UBER 在臺營運取締與抗議話題之三元決策情感趨勢圖。

3. 稅務與收費

稅務與收費議題是 UBER 在臺營運討論文章量排名第 3 的話題，達 53 篇，其中正面

文本 19 篇、中立文本 10 篇和負面文本 26 篇，整體情感傾向偏向負面，多針對 UBER 在臺營運尚未具備與計程車產業繳稅的相同基礎為主，多提及繳稅、規避稅捐、逃稅、車資、營利事業所得稅等關鍵字。UBER 未來策略走向應朝繳稅與計程車制度調整進行，否則形同於不公平的市場競爭，畢竟 UBER 服務對司機的高抽佣以及節省的稅捐成本，將對國內計程車市場造成極大的打擊。政府亦倡導共享經濟在臺發展之重要性，若 UBER 在臺營運合法，則有催生創新服務產業的意義，因此 UBER 須先取得合法執業許可及繳稅，方可在臺營運。

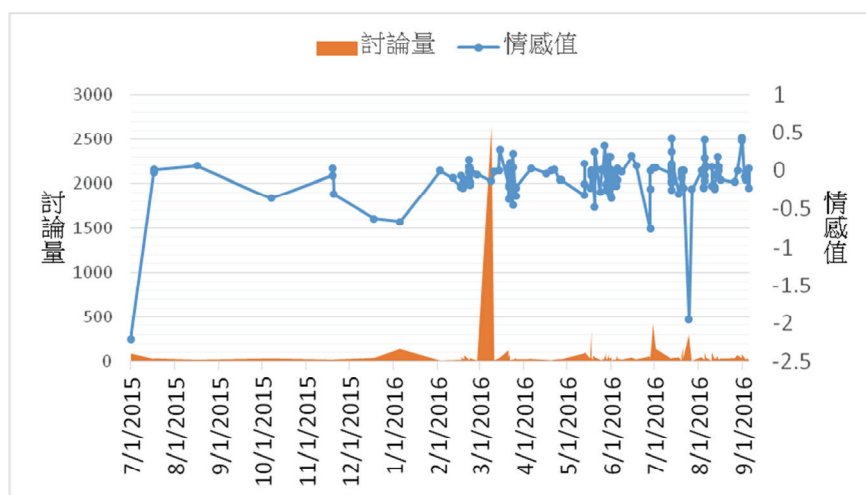


圖 5 UBER 在臺取締與抗議話題之三元決策情感趨勢圖

五、結論與建議

5.1 結論

1. 本研究回顧中文文本分類及情感分析研究之相關文獻，發現以二元分類作為情感分析之類別，應用較廣。但近年來提高文本自動分類準確率的三元決策理論則愈受重視。因此本研究建構三元決策之文本情感分析模式，以 UBER 在臺營運話題之網路文本為實證對象。實證分析結果顯示可有效將文本劃分為正面、中立與負面之情感傾向類別，進而區分不同時期 UBER 各營運話題之情感傾向。
2. 藉由詞頻權重法計算篇章級文本之情感傾向，可綜整出 UBER 在臺營運話題之情感趨勢。自 2015 年 7 月至 2016 年 9 月間，民眾對於 UBER 在臺營運之共乘平台服務的評價，整體情感趨勢係以正面偏中立為主，高於負面情感傾向。期間產生 4 次討論爆量之輿情尖峰事件（討論量大於 1,500 則回文），主要話題包含營運制度、取締與抗議、稅務與收費。進一步分析得知，民眾對於 UBER 在臺營運話題之整體情感傾向，以正面偏中立為

主。尤其對於 UBER 服務內容及品質則最為讚許，討論量依序為動態調整費率、司機評價機制、圖資定位資訊應用、優質使用介面與網路媒合、清楚的導航路線、回應顧客需求迅速等項目，此應可作為政府推動多元計程車方案的政策參考。

3. 在整體負面趨勢情感趨勢則在 UBER 營運合法性之爭議，討論量以非法營業與未盡繳稅義務最多，顯示目前 UBER 在臺並無商業登記以及繳交營利所得稅，是 UBER 在臺營運首要解決的問題。
4. 運用三元決策理論為基礎的文本情感傾向分類，以重心法解模糊數計算出中立門檻值介於-0.0225 與-0.3196 之間，可將原本二元分類定義下為輕微負面的文本劃分至中立區間，其中正面文本 374 篇、中立文本 307 篇以及負面文本 117 篇。在負面文本中提及法令、法規制度、制定專法等關鍵字數量最多，顯示 UBER 的爭議始終集中於在臺營業之合法性。

5.2 建議

1. 建議後續可運用三元分類結果，將關鍵詞頻統計結合單篇的網路文本，建構符合 UBER 在臺營運的情感認知迴歸模式，區分不同營運話題之新增文本，是否對 UBER 營運的情感認知之變化有顯著差異。
2. 本研究利用關鍵情感詞庫之詞頻統計方法來計算特定研究領域情感值，其社群網路本的情感值將受到文本母體總數量增減，而使單一情感特徵值的數值產生變化，因此本研究三元決策之分類門檻值計算，係以將近 1 年時間週期 (2015.07~2016.09) 的有效樣本為基礎，始得出該年度之通用門檻值。建議不同年度需重新彙整蒐集新增之意見文本，重新計算該研究領域的通用情感值，調整至最佳文本分類效果。

參考文獻

1. Cramer, J. and Krueger A., “Disruptive Change in the Taxi Business: The Case of Uber”, *American Economic Review*, Vol. 106, No. 5, 2016, pp. 177-182.
2. Zafarani, R., Ali Abbasi, M., Liu, H., *Social Media Mining: An Introduction*, 1st Ed., Cambridge University Press, UK, 2014.
3. 劉知遠等人編著，**大數據智能**，第 1 版，電子工業出版社，北京，2016。
4. Drury, G., “Opinion Piece: Social Media: Should Marketers Engage and How Can It Be Done Effectively”, *Journal of Direct, Data and Digital Marketing Practice*, 2008, pp. 274-277.
5. Manovich, L., “Trending: The Promises and the Challenges of Big Social Data”, in M. K. Gold (Ed.), *Debates in the Digital Humanitie*, http://www.manovich.net/DOCS/Manovich_trending_paper.pdf, The University of Minnesota Press, Minneapolis, MN, 2011.
6. Chan, S. M., Cho, M., and Lee, S., “User Perceptions of Social Media: A Comparative Study

- of Perceived Characteristics and User Profiles by Social Media”, *Journal of Communication and Media Technologies*, 2013, pp. 149-178.
7. Wilson, K., Fornasier, S., and White, K. M., “Psychological Predictors of Young Adults' Use of Social Networking Sites”, *Cyberpsychology, Behavior and Social Networking*, Vol. 13, 2013, pp. 173-177.
 8. 朱斌妤、黃東益、洪永泰、李仲彬、曾憲立，「數位國家治理(2)：國情追蹤與方法整合」，電子治理研究中心委託研究報告，民國 104 年。
 9. 蕭乃沂、陳敦源、廖洲棚，「政府應用巨量資料精進公共服務與政策分析之可行性研究」，電子治理研究中心委託研究報告，民國 104 年。
 10. 陶冶中、陳亭愷，「社群運算應用於網路輿情感傾向分析之研究－以實施國道計程電子收費政策為例」，*運輸學刊*，第 28 卷，第 3 期，民國 105 年，頁 295-334。
 11. Das, S. and Chen, M., “Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards”, *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, 2001, pp. 1367-1373.
 12. Dave, K., Lawrence, S., and Pennock, D. M., “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews”, *Proceedings of WWW*, 2003, pp. 519-528.
 13. Pang, B. and Lee, L., “Opinion Mining and Sentiment Analysis”, *Foundations and Trends in Information Retrieval*, Vol. 2, No.1-2, 2008, pp. 1-135.
 14. Liu, B., “Sentiment Analysis and Opinion Mining”, *Synthesis Lectures on Human Language Technologies*, Vol.5, No.1, 2012, pp. 1-167.
 15. 許鑫，*基於文本特徵計算的信息分析方法*，第 1 版，上海科學技術文獻出版社，上海，2015。
 16. Hatzivassiloglou, V. and McKeown, K. R., “Predicting the Semantic Orientation of Adjectives”, <http://www.aclweb.org/anthology/P97-1023>, 2016.
 17. Wiebe, J. M., “Learning Subjective Adjectives from Corpora, Department of Computer Science”, <http://people.cs.pitt.edu/~wiebe/pubs/papers/aaai2000.pdf>, New Mexico State University, 2016.
 18. Riloff, E. and Wieber, J. M., “Learning Extraction Patterns for Subjective Expressions”, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, 2003, pp. 105-112.
 19. Turney, P. D., “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Review”, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002, pp. 417-424.
 20. Hu, M. and Liu, B., “Mining and Summarizing Customer Reviews”, *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2004)*, 2004, pp. 168-177.
 21. 黃心宜，「基於影響力分析之意見單元評價的研究」，淡江大學資訊管理學系碩士論文，民國 103 年。
 22. 楊惠淳，「以主客觀分析與相互資訊檢索探討情感分析之準確度－以電影評論為例」，臺北科技大學資訊與運籌管理研究所碩士論文，民國 100 年。

23. Pang, B., Lee, L., and Vaithyanathan, S., “Thumbs Up? Sentiment Classification Using Machine Learning Techniques”, Proceedings of EMNLP '02 of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Vol. 10, 2002, pp. 79-86.
24. Hofmann, T., “Probabilistic Latent Semantic Indexing”, Proceedings of the Conference on Uncertainty in Artificial Intelligence, 1999, pp. 1367-1373.
25. Blei, D. M., Andrew, Y. Ng., and Michael I. J., “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, Vol. 3, 2003, pp. 993-1022.
26. Zhao, W.X., Jiang J., Yan, H., and Li, X., “Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid”, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2010, pp. 56-65.
27. Sauper, C., Haghighi, A., and Barzilay, R., “Content Models with Attitude”, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011, pp. 350-358.
28. Arjun, M. and Liu, B., “Aspect Extraction through Semi-Supervised Modeling”, Proceedings of the 50th Annual Meeting of Association for Computational Linguistics, <https://www.cs.uic.edu/~liub/publications/ACL-2012-aspect-extraction.pdf>, 2012.
29. Go, A., Bhayani, R., and Hung, L., “Twitter Sentiment Classification Using Distant Supervision”, CS224N Project Report, Stanford, 2009, pp. 1-12.
30. Pak, A. and Paroubek, P., “Twitter as a Corpus for Sentiment Analysis and Opinion Mining”, *LREC*, 2010, pp. 1320-1326.
31. Bermingham, A. and Smeaton, A. F., “Classifying Sentiment in Microblogs : Is Brevity an Advantage ? ”, Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, 2010, pp. 1833-1836.
32. Johan, B., Gonçalves, B., Ruan, G., and Mao, H., “Happiness is Assortative in Online Social Networks”, *Artificial Life*, Vol. 17, No. 3, 2011, pp. 237-251.
33. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., and Li, P., “User-Level Sentiment Analysis Incorporating Social Networks”, Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 1397-1405.
34. Liu, B., Hu, M., and Cheng, J., “Opinion Observer : Analyzing and Comparing Opinions on the Web”, Proceedings of the 14th International Conference on World Wide Web, ACM, 2005, pp. 342-351.
35. Lu, Y., Duan, H., Wang, H., and Zhai, C., “Exploiting Structured Ontology to Organize Scattered Online Opinions”, Proceedings of International Conference on Computational Linguistics, <http://aclweb.org/anthology/C10-1083>, 2010.
36. Yang, H., Si, L., and Callan, J., “Knowledge Transfer and Opinion Detection in the TREC 2006 Blog Track”, Proceedings of TREC, <http://www.cs.cmu.edu/~callan/Papers/trec06-huiyang.pdf>, 2006.
37. Blitzer, J., Dredze, M., and Pereira, F., “Biographies, Bollywood, Boomboxes and Blenders: Domain Adaptation for Sentiment Classification”, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 2007, pp. 432-439.
38. Yao, Y. Y., “An Outline of a Theory of Three-Way Decisions”, Proceedings of the 8th the

International RSCTC Conference, 2012, pp. 1-17.

39. 劉盾、李天瑞、李華雄，「粗糙集理論：基於三支決策視角」，*南京大學學報*，第 49 卷，第 5 期，2015 年，頁 574-581。
40. 田海龍、朱豔輝、梁韜、馬進，「基於三支決策的中文微博觀點句識別研究」，*山東大學學報(理學版)*，第 49 卷，第 8 期，2014 年，頁 59-65。
41. 張志飛、王睿智、苗奪謙，「第七章：基於三元決策的多粒度文本情感分類」，*三支決策：複雜問題求解方法與實踐*，第 1 版，科學出版社，北京，2015 年。
42. 王磊、黃河笑、吳兵、鄭任兒，「基於主題與三支決策的文本情感分析」，*計算機科學*，第 42 卷，第 6 期，2015 年，頁 93-96。
43. 周哲、商琳，「一種基於動態詞典和三支決策的情感分析方法」，*山東大學學報(工學版)*，第 45 卷，第 1 期，2015 年，頁 19-23。
44. Luhn, H. P., "The Automatic Creation of Literature Abstracts", *IBM Journal of Research and Development*, No. 2, 1958, pp. 159-165.
45. Pawlak, Z., "Rough Sets", *International Journal of Computer and Information Sciences*, No. 11, 1982, pp. 341-356.