

國立政治大學經濟學(系)研究所

碩士學位論文

大數據分析在社群網站的應用—以影響PS4銷

售量因素為例



指導教授：王信實 博士

研究生：蔡仲銘 撰

中華民國 107 年 7 月

摘要

隨著網路與行動裝置的蓬勃發展，人們使用社群網站的比例快速增加，廠商也開始重視社群網路裡的廣告活動與粉絲經營。研究指出，消費者在購買商品前後會到社群網站中搜尋相關評價與發表心得的比例也逐漸上升。本研究主要以電玩遊戲主機 PlayStation 4(ps4)為例，利用 R 軟體蒐集推特(Twitter)上的北美地區推文(tweets)大數據資料與 Google Trend 搜尋熱度資料進行統計分析，建立支援向量迴歸模型(SVR)進行商品銷售量的預測並與其他模型比較。本研究發現，社群網站上的每周推文數、正面情感推文占總推文數的比例以及 Google Trend 的搜尋熱度指標越高會對商品的銷售量有正面影響，而文章包含影音圖片等外部 url 連結的廣告效果會比純文字來的更好。此外，當推特上的推文顯示消費者處於愛達(AIDA)模式中表現期的推文比例越高，也與商品的銷售量有正相關。另外，透過字詞分析也發現，遊戲廠商經常會舉辦抽獎活動或促銷活動作為增加官方推特的互動率、培養粉絲以及增加廣告文章的曝光率的手段。遊戲產業在社群網站上的廣告策略應以影音圖片為主、文字為輔，並以擴散情報及玩家間的正面討論為首要目標。

關鍵字:大數據、社群網站、Twitter、SVR、AIDA 模型

目次

第一章 緒論.....	1
第一節 研究背景.....	1
第二節 研究動機.....	2
第三節 研究目的.....	3
第四節 研究架構.....	5
第二章 產業背景.....	7
第三章 文獻探討.....	17
第一節 社群網站相關研究.....	17
第二節 AIDA模式.....	26
第三節 大數據相關研究.....	29
第四章 資料說明.....	35
第一節 研究對象.....	35
第二節 資料收集.....	36
第三節 資料概述.....	38
第四節 變數說明.....	39
第五章 研究方法.....	43
第一節 模型介紹.....	43
第二節 預測指標.....	46

第六章 研究結果.....	48
第一節 Twitter數據敘述統計與每週趨勢.....	48
第二節 字詞分析與文字雲.....	54
第三節 預測模型.....	58
第七章 結論與建議.....	64
第一節 結論.....	64
第二節 建議、研究限制與未來研究方向.....	66
參考文獻.....	68



表次

表1 遊戲裝置歷史銷售量前十名.....	8
表2 三大社群網站比較表.....	15
表3 社群網站研究文獻回顧.....	23
表4 AIDA與HoE模式比較表.....	28
表5 大數據研究文獻回顧.....	33
表6 蒐集資料方式比較表.....	38
表7 本研究資料蒐集概況.....	39
表8 變數說明.....	42
表9 MAPE數值區間.....	47
表10 敘述統計表.....	49
表11 詞頻統計表.....	58
表12 OLS模型估計結果.....	60
表13 SVR與OLS預測效率比較表.....	63

圖次

圖1 ICT應用服務認知度.....	2
圖2 PS4、Xbox One以及Wii U累計銷售量.....	10
圖3 2017~2018年各品牌VR裝置出貨量.....	11
圖4 各大社群網站月活躍用戶比較.....	12
圖5 三大社群網站情報擴散速度比較.....	16
圖6 國人擁有社群帳號比例.....	19
圖7 AISDAS循環示意圖.....	27
圖8 Twitter推文屬性.....	35
圖9 twitter API主介面.....	36
圖10 Google Trend示意圖.....	40
圖11 SVM平面示意圖.....	43
圖12 Twitter推文數每週趨勢.....	50
圖13 Twitter推文包含url比例的每週趨勢.....	51
圖14 Twitter推文PNratio每週趨勢.....	51
圖15 屬於認知期的推文比例每週趨勢.....	52
圖16 屬於情感期的推文比例每週趨勢.....	53
圖17 屬於行為期的推文比例每週趨勢.....	53
圖18 美國地區ps4每週銷售量趨勢.....	54

圖19 Twitter推文內容文字雲..... 56

圖20 SVR模型的PS4銷售量的樣本內預測數值與實際數值折
線圖..... 62



第一章、緒論

第一節 研究背景

隨著科技的日新月異，以及新產品的推陳出新，人們的生活習慣與休閒娛樂方式也隨著改變了。其中最關鍵的突破便是網際網路的出現，每天上網已經成為台灣民眾不可或缺的休閒娛樂之一，根據財團法人台灣網路資訊中心（Taiwan Network Information Center, TWNIC）公布的 2017 台灣寬頻網路使用調查報告，台灣整體的上網率達到 80%，其中 39.1% 是使用行動網路上網，顯示在行動載具普及之後，民眾的上網方式也逐漸從個人電腦轉移到智慧型手機、平板。在上網的年齡層方面，根據 TWNIC 公布，以 18-30 歲的年齡層最高，成為我國網路主要使用族群。

另外根據財團法人資訊工業策進會 (Foreseeing Innovative New Digiservices, FIND) 針對民眾創新科技生活應用服務包括健康照護、生活環境、交通、政府、育、樂、購物消費、數位溝通等八大構面進行調查所公布的「2014 年台灣民眾之智慧化生活有感調查」(吳佩玲, 2015)，文章中討論台灣人民對於網際網路運用於消費、通訊的感受與滿意程度調查。文中結果顯示：台灣使用者在使用度方面，最常使用功能的前五名依序為「手機網路通訊 (97.3%)」、「智慧交通付費 (96.8%)」、「數位影音 (96.3%)」、「數位交易 (95.3%)」，以及「社交網路 (93.8%)」，以上皆超過 9 成 (圖 1)。其中，「手機網路通訊」包含了收發簡訊、郵件，以及即時通訊，而「社交網路」包含部落格與社群網站等，這兩項服務在台灣都已有相當高的知名度，其中目前幾個較為知名的社群網站分別為 Facebook、Instagram，以及 Twitter。Facebook 成立於 2004 年，目前在全球已經擁有超過 20 億使用者，是世界上最大的社群網站。Instagram 成立於 2010 年，目前有 6 億用戶，排行第

二。而 twitter 成立於 2006 年，有 3.3 億用戶，其中絕大部分的社群網站使用者都是透過行動網路來使用 (Global Digital Statshot, 2017)。「數位影音」認知度也很高 (96.3%)，表示民眾對於聯網電視、線上影視、線上遊戲以及線上音樂/廣播或數位音樂、鈴聲有一定程度的了解。顯示人們在目前利用網路進行行動通訊及社交互動這一部分已經十分普及。

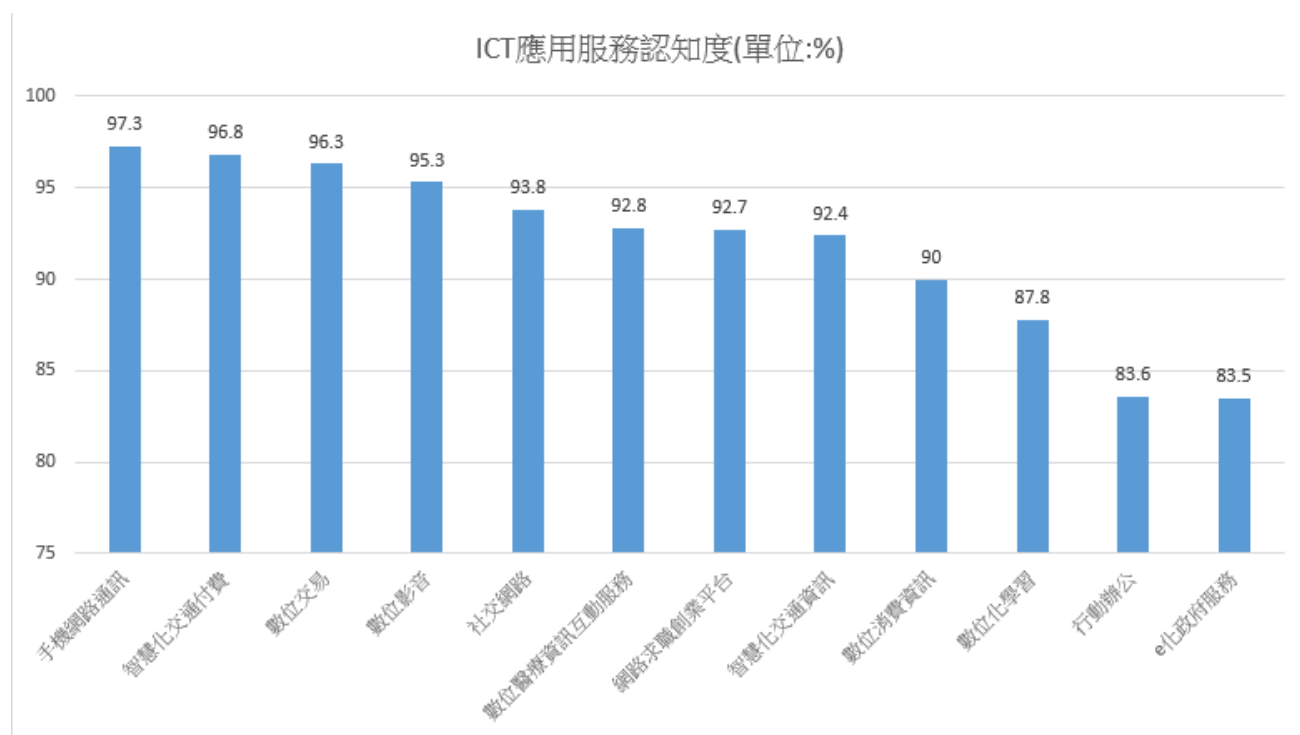


圖 1:ICT 應用服務認知度

資料來源: 資策會創新應用服務研究所 FIND

第二節 研究動機

PlayStation 4 (ps4) 是目前最暢銷的電視遊樂器，加上近幾年 VR 裝置崛起，被視為未來相當有潛力的產業，而其中一大應用即是將 VR 裝置與遊戲機結合，

創造全新的遊戲體驗。雖然 ps4 主機在各大討論版或是社群網站都有許多分析、評論這台主機優缺點的文章，以及來自眾多使用者第一手的遊玩心得等，這個領域也有一些先前做過的相關研究。但傳統的量化研究方法因受限於母體過大或母體未知等因素，往往採用「隨機抽樣」或是「限定某一研究區間」的方式，以「樣本推論母體」的統計方式對其進行分析；然而隨著科技逐漸的發展進步，誕生出許多新的資料蒐集與統計工具可以解決資料量過大而無法進行分析的問題，大數據分析的核心在於「分析全部或盡量完整的資料」，即以未經過抽樣的數據探討資料中的細節，大量的數據分析，除了能夠使研究者更全觀的探討研究對象的樣貌外，同時也能解決傳統統計方法存在資料偏差的問題 (Hargittai, 2015)，並發掘出傳統量化分析方法無法得知的現象 (Schonberger & Cukier, 2013/林俊宏譯，2013；城田真琴，2012/鐘惠真和梁世英譯，2013)。而且在傳統的專業討論版中，玩家受到「板規」的限制，例如發文字數限制或是受到討論版的立場影響，無法盡情抒發個人真實的意見。然而在社群網站上，個人帳號發表的文章幾乎不會受到他人限制，可以發表個人對某商品最真實的看法，因此本研究選擇以社群網站上的文章作為主要研究對象。而在 Lassen et al. (2014) 的研究中，他們認為發出推文 (tweets) 的舉動可以視為是使用者對於某個商品感興趣而想深入了解或是評論它，因此才會發表一則推文。另外在 Asur & Huberman 的研究中也運用了社群網站的資料成功地來預測電影的票房。本研究主要利用大數據分析，蒐集來自社群網站推特 (Twitter) 的大數據資料進行分析，並探討是否這些大數據資料能運用於預測 ps4 的銷售量。

第三節 研究目的

在研究背景與研究動機中提到，隨著科技快速發展，民眾蒐集情報的方法也更加及時、有效率。以往想要購買某種商品時，通常會先詢問親朋好友使用心得

或是透過報章雜誌吸收資訊。但在網際網路出現之後，這些蒐集的行為通通轉移到網路上的討論區或是 BBS 站等專業討論版。近幾年行動網路與社群網站的崛起又帶動一波情報來源的轉移，大多數人選擇在社群網站上分享自己對某商品的使用心得、優缺點及評價。不只是消費者，就連廠商也進入社群平台成立粉絲專業與消費者互動、溝通，或是買廣告來吸引消費者。Web 2.0 的概念讓我們不再只是利用網際網路單方面吸收訊息，同時能夠在社群平台上進行多方面交流互動與討論，社群媒體的概念也隨之而生。因此，網路上輿論的力量也逐漸受到重視。過去媒體作為資訊的傳播者，有著時空傳遞的限制，現在資訊取得變得更多元，尤其資訊在透過網友反覆的討論與交流，逐漸形成一股輿論力量（謝元晟等，2016）。

本研究希望透過大數據分析結合傳統統計分析與文字探勘技術，並以 Elmo St. Lewis 最初建立的愛達 AIDA (Awareness, Interest, Desire, Action) 模型與 Lavidge & Steiner 的 Hierarchy of Effects (HoE) 模型為理論基礎，針對 ps4 的銷售量做分析，並以支援向量迴歸 (Support Vector Regression, SVR) 建立模型後進行預測與回答以下問題：

- 一、PS4 的銷量趨勢變化情形，是否與社群網站上的發文聲量的變化有關？
- 二、某一段時間內的 twitter 用戶對 ps4 的討論的情感分數高低是否將反映在當期 ps4 的銷售量上？
- 三、根據 AIDA 模型中的消費三階段，在蒐集資料的過程中，twitter 用戶位於這三階段的比例是否改變？
- 四、官方經營 twitter 專頁的積極度是否能影響社群網站中潛在消費者的購買意願？

五、搜尋引擎的搜尋熱度變化是否也能反映在當期 ps4 的銷售量變化上?

根據余峰璋的文章中提到社群網站的相關學術研究大多只包含以閱聽人為主的內容，也就是透過使用者的觀點來探討社群網站相關議題的成效與影響。即使有部分研究以企業角度切入但也僅止於廣告的投放與內部運作等方面，而對於粉絲專業的經營管理與內容與內容呈現等實務操作較少涉及（余峰璋，2017）。因此，本研究將官方粉絲專業的經營也納入分析，除了考量 ps4 官方推特的粉絲數之外，也包含官方推文的情感分析分數與互動率，另外也包含官方對 ps4 主機與遊戲的宣傳力度（以包含外部連結 url 的文章數佔總文章數的比例呈現）對主機銷售量的影響等，希望納入官方推特經營的角度，呈現各個面向的社群網站對產品銷售量影響的分析。

最後，希望透過迴歸分析、字詞分析，與文字雲等分析方法的結果找到社群網站中的那些數據能確實反映在遊戲商品的銷售量上，如官方經營粉絲團的方式、發表的文章類型，或是一般使用者正面情感數值的變動等因素是否可以用來預測產品的銷售量等。將各種因素總結之後，建立一個迴歸模型用於預測遊戲產業相關產品的銷售量，並提供遊戲產業如何在社群網站上經營粉絲專頁，強化廣告效果的參考。

第四節 研究架構

本文章的研究架構設計如下：第二章為產業背景，文中將探討電動遊戲主機 ps4 目前發展的概況、VR 產業的潛力與未來應用，以及介紹各社群網站的功能差異、經營情況與活躍人數等內容。第三章為文獻探討，該章節包含以社群網站經營與廠商經營粉絲團策略為研究的文章、以社群網站使用者為對象的大數據研究，以及 AIDA 模型的發展研究等三大項。第四章為資料說明，其中包含利用 R

軟體蒐集到的社群網站資料的屬性解釋、本研究蒐集到的社群網站資料概述，以及本研究使用的變數說明與測量方法。第五章為模型介紹，本研究使用機器學習中常用的支援向量迴歸 (SVR) 作為研究模型，在此章節中會介紹模型的數學設定以及用於模型預測的預測指標的公式。第六章是研究結果，其中有敘述統計趨勢、字詞統計與文字雲、模型結果與預測能力等三個部分。最後第七章為結論、研究限制與未來建議。



第二章 產業背景

隨著網路發展越來越進步，速度提升與連線品質的穩定使許多遊戲都開始提供線上多人遊玩的選擇。不只是電腦線上遊戲，家用主機如 playstation、xbox 等系列主機也開始重視連線功能並建立專屬的線上平台。其中 sony 的 playstation 最新系列的家用主機 ps4 在短短 4 年內銷量已經突破 7000 萬部，不只壓倒性的贏過其他競爭對手的同期主機 xbox one (3600 萬部)、wii U (1350 萬部)，甚至還超越了史上最暢銷的主機 ps2 的同期銷量 (VGChartz, 2017)。Sony 不只在家用主機上取得勝利，在近幾年逐漸發展的「虛擬實境(VR)」產業也有相當出色的成績。根據國際研究顧問機構 Gartner 在 2017 年的調查顯示，頭戴式顯示器佔該年所有穿戴式裝置出貨數的 7%，出貨量約 2200 萬部，主要是以遊戲玩家居多 (Gartner, 2017)。另外根據 Sony 最新報告，旗下的 VR 裝置 Playstation VR (PSVR) 自 2016 年 10 月發售以來，截至 2017 年 12 月 3 日在全球已經超過 200 萬台銷售量，而家用主機 PS4 則總共銷售出 7060 萬台。這也讓 PSVR 成為市場中銷售成績最好的 VR 裝置 (SONY, 2017)。根據研究調查機構 Canalys 在 2017 年第三季統計，目前 PSVR 在市場中擁有約 49% 市佔，再來是 Oculus Rift 的 21%，第三則為 HTC Vive 的 16%。這些數據顯示了 ps4 不僅在目前已成為同期最暢銷的家用遊戲主機，在未來還能搭上潛力無窮的 VR 遊戲產業，使銷量繼續上升。下表 1 為遊戲裝置的歷史銷量排行，可以看到 ps4 位於第 10 名的位置，一到九名都是產品生命週期結束的主機，而 ps4 約還有 5 年左右。因此 ps4 能非常有機會能超越前九名的主機，登上史上最暢銷的主機第一名。

表 1:遊戲裝置歷史銷售量前十名 (2018)

排名	主機	北美	歐洲	日本	其他地區	全球
1	PlayStation 2 (ps2)	53.65	55.28	23.18	25.57	157.68
2	Nintendo DS (NDS)	57.39	52.07	33.01	12.43	154.90
3	Game Boy (GB)	43.18	40.05	32.47	2.99	118.69
4	PlayStation (PS)	38.94	36.91	19.36	9.04	104.25
5	WII	45.51	33.88	12.77	9.48	101.64
6	PlayStaion 3 (ps3)	29.42	34.55	10.47	12.46	86.90
7	Xbox 360 (X360)	49.11	25.87	1.66	9.16	85.80
8	Game Boy Adavance (GBA)	40.39	21.31	16.96	2.85	81.51
9	PlayStation Portable (PSP)	21.41	24.14	20.01	15.26	80.82
10	PlayStation 4 (PS4)	25.20	29.62	5.95	13.24	74.01

單位:百萬台

資料來源:VGChartz(2017), <http://www.vgchartz.com/platforms/>

自從 2000 年發售的 PlayStation 2 (PS2) 以超過 1 億 5000 萬部的銷售量寫下歷史紀錄後，到目前為止都沒有其他遊戲裝置能超越這個紀錄。再加上智慧型手機、平板的出現，讓遊戲主機的版圖受到侵蝕，一般認為在手機、平板出現之後電視遊樂器的市場便逐漸萎縮了 (statista, 2014)。但根據國外網站 VGChartz 所公布的電視遊樂器銷售數據來看，在智慧型手機崛起的幾年(2012~2015)內，電視遊樂器的銷售量整體雖然呈現下滑趨勢，整體遊樂器市場銷售量從 2012 年的約 950 萬台下降至 2015 年的約 700 萬台，但在 SONY 以及微軟都推出新一代的主機 (PS4, XBOX ONE) 之後，銷售量又有提升的趨勢 (VGChartz, 2015)。

然而，在 PS4 推出之後，不只以超過 50% 的市佔率稱霸同世代的遊戲主機 (IHS Markit, 2017)，甚至超過歷史銷售冠軍 PS2 的同期銷售量。再加上，2017 年 3 月任天堂推出的 Nintendo Switch (NS) 一上市僅花了 9 個月就賣出超過 1000 萬台。而在 2018 年 1 月底公開的第三季 (截止至 2017 年 12 月底) 任天堂財報顯示，任天堂 Switch 主機已在全球售出 1,486 萬台，也超越了 PS4 上市首年的銷售紀錄 (1,440 萬台)。原本在前幾年被預期會逐漸萎縮的遊樂器市場似乎又出現了起死回生轉機。

另一個對遊樂器市場復甦有利的關鍵在於虛擬實境 (Virtual Reality, VR) 產業。虛擬實境也稱虛擬環境、虛擬技術，是利用電腦模擬一個三維空間的虛擬世界，提供使用者視覺等感官的模擬，讓使用者感覺彷彿身歷其境，可以及時、沒有限制地查看三維空間內的事物。當使用者進行位置移動時，電腦可以立即進行複雜的運算，將精確的三維影像傳回產生臨場感。這項技術整合了電腦圖形、電腦仿真、人工智慧、感應、顯示及網路並列處理等技術的最新發展成果，是一種由電腦技術輔助生成的高技術模擬系統 (Wikipedia)。2016 年被稱作「虛擬實境 (VR) 元年」，雖然 VR 的大眾市場尚未成熟，根據國際研究顧問機構 Gartner 調

查顯示，頭戴式顯示器佔該年所有穿戴式裝置出貨數的 7%，出貨量約 2200 萬部，但已經逐漸運用在遊戲方面了，例如 VR Arcade（街機）模式。簡單來說，它的營運模式類似現在的電玩娛樂場所，將 VR 遊戲與機台結合後，可以讓對新科技感到好奇的消費者們，以單次消費的模式、較為平實的價格實際使用 VR（數位時代，2017）。另外像是 HTC、Sony、Oculus，以及微軟都有研發 VR 裝置運用在電玩上。2017 年 Sony 在 PS4 Pro 熱銷以及大型開發商的支持下，旗下 VR 裝置 PS VR 出貨量達到 170 萬台，市占率最大，而 Oculus Rift 與 HTC Vive 則分別以 70 萬台與 50 萬台緊追其後（TechNews, 2018）。專家也推測，未來 10 年內，VR 最被看好的兩個領域為遊戲與電影（STOCKFEEL, 2017）。因此，相信在未來，各大家用遊戲主機將陸續成為 VR 裝置的發展應用平台。

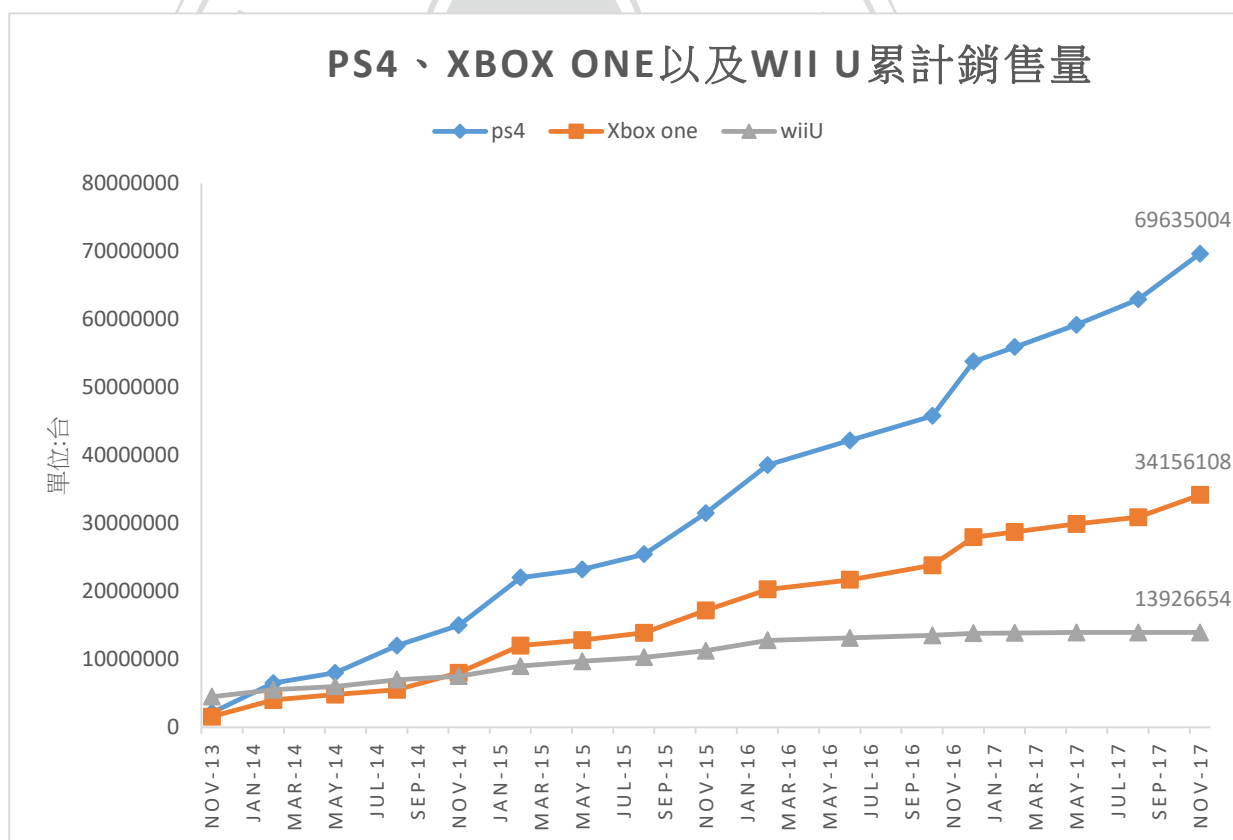


圖 2:PS4、Xbox One 以及 Wii U 累計銷售量

資料來源:VGChartz (2017) , <http://www.vgchartz.com/article/268377/ps4-vs-xbox-one-vs-wii-u-global-lifetime-salesapril-2017-update/>

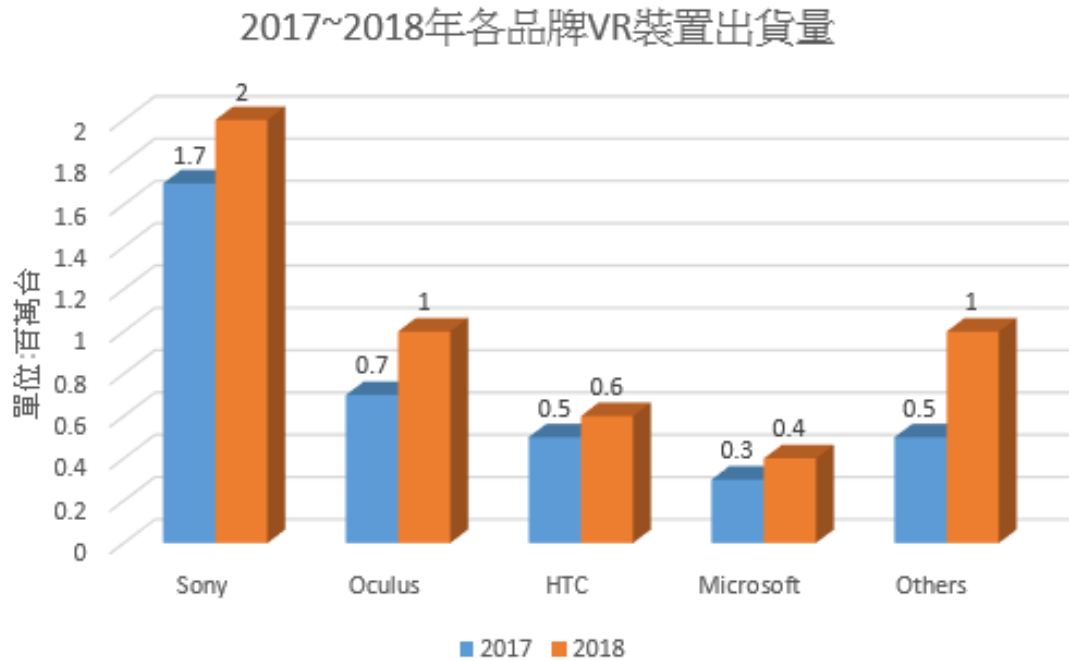


圖 3:2017~2018 年各品牌 VR 裝置出貨量

資料來源:拓璞產業研究院 (2018)

另外，根據資策會 (2011) 網路社群媒體發展分析報告，台灣民眾在買東西前會上網蒐集資料、使用心得以及評價的情報的比例超過 70%。同時，Google 公布最新《消費者洞察報告》及《台灣數位消費者研究報告》，公開台灣消費者最新的上網與購物決策行為，其中有 69% 的台灣消費者在購物之前會先上網研究，更有 76% 的人會每天利用網路蒐集資訊 (Google, 2016)。在購買涉入度較高且未使用無法深入了解產品特性的商品 (例如:3C 商品) 時，更會傾向先蒐集相關資料，再決定是否購買。而這些心得、評價在行動網路與社群平台蓬勃發展的今日，

大多來自於部落格或 YouTube 名人、相關的大型討論區，以及 Facebook、Twitter 等社群網站的情報交換 (林怡靜，2015)。

現在的社群網站五花八門，那些社群網站才是大多數人較頻繁使用的選擇？根據 techcrunch 在 2017 年做的調查，Facebook 的月活躍用戶已經突破 20 億，高居第一，而 YouTube 以 15 億月活躍用戶量位居第二。另外還有 WhatsApp 和 Facebook Messenger 兩大通訊軟體超過 10 億月活躍用戶，兩者都是 12 億。

除了以上幾個已經突破 10 億月活躍用戶大關的社群軟體之外，微信在 2016 年達到的 8.89 億；Instagram 的 7 億月活躍用戶，以及 Twitter 的 3.28 億都是在全世界名列前茅的社群網站 (techcrunch, 2017)。

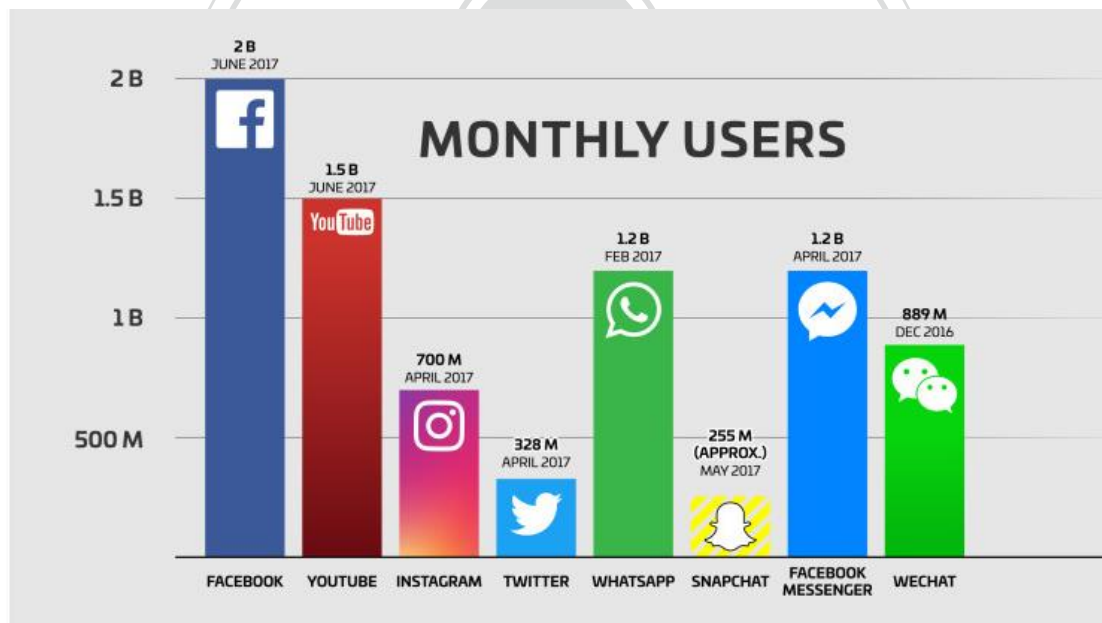


圖 4: 各大社群網站月活躍用戶比較

資料來源: techcrunch

Twitter (官方中文譯名推特) 是一個社群網路與微網誌服務，它可以讓用戶更新不超過 140 個字元的訊息，這些訊息也被稱作「推文 (Tweet)」。推特成立於 2006 年，截至 2017 年，Twitter 共有 3.28 億月活躍用戶。Twitter 被形容為「國際網路的簡訊服務」。網站的非註冊用戶可以閱讀公開的推文，而註冊用戶則可以通過 Twitter 網站、簡訊或者各種各樣的應用軟體來發布訊息 (維基百科)。Twitter 跟其他社群網路所不同的是，Twitter 是以興趣網路為主所形成的集合，使用者可以在 Twitter 上面關注所感興趣的議題，並且具有即時、公開、對話、分享等四大特質，使用者可以透過 Twitter 瞭解世界各地發生的重大事件，並且參與其中，這是其他社群網路所難以辦到的事情 (Greenberger, 2015)。同時，Twitter 還被認為在阿拉伯之春、以及 2010 年代後期以來全球的政治活動中扮演重要角色。例如：阿拉伯之春的埃及各國民眾透過社交媒體集體守門與框架來界定 #egypt 標籤的相關討論議題 (Meraz & Papacharissi, 2013)。

雖然 Twitter 在台灣的使用率偏低，約 5%，但在美國約有 3 成的青少年會使用 Twitter，其中 18 到 29 歲族群有將近四成是推特用戶，北美也佔了推特收入的 65% (PewReserchCenter, 2015)。在日本，Twitter 用戶數更是超越了 Facebook，根據 Twitter 在 2016 年首次公開的該公司的海外市場用戶數量，Twitter 在日本的月活躍用戶數高達 3,500 萬，比起 Facebook 同期月活躍用戶數高出 1,000 萬。另外就 Twitter 全球用戶發文使用的各國語言的比例而言，根據 SemioCast 在 2013 年的報告數據顯示，全球推文使用語言比例最高的是英文占 34%，其次為日文占 16%，第三名為西班牙語占 12%，中文推文則未進入排行前十名的語言；另一份 GNIP 的調查數據顯示，在 2012 年 Twitter 使用的語言中，簡體中文排名第 17、繁體中文排名第 19，兩種中文語言的推文量皆不到全世界推文量的 1% (Lehman, 2014; MIT Technology Review, 2013)。

Twitter 做為社群網站有幾個不同於其他社群網站的優點:

一、推文預設公開:有別於 Facebook 的使用者通常將訊息設定為隱私或只有好友才能看到, Twitter 的推文預設為公開, 因此所有人都能看到。與 Facebook 不同, Twitter 的情報散播速度很快, 其原因來自 Twitter 用戶在散播情報時, 是以自己為中心向所有其他用戶發散。

二、建立情報網:由於在 Twitter 上情報散播速度很快, 蒐集使用者本身有興趣的話題情報自然是多數人的目的。只要用戶追隨 (follow) 自己想關注的用戶便可以透過本身頁面的時間軸即時看到各種情報, 能用秒為單位來吸收訊息, 而時間軸排序也不會讓情報的先後順序錯亂。

三、追隨者而非朋友關係:Twitter 中並沒有雙方成為朋友的概念, 只需要單方面追隨就可以在自己的頁面看到對方的推文, 但在對方沒有追隨自己的情況下, 對方就不會看到自己的推文。而追隨的動作是不需要經過對方確認, 與 Facebook 的好友不同點是追隨是單向而好友是雙向的概念。

四、無地域性:以微博來說, 使用者大多是中國用戶, 區域性明顯, 又因為微博有言論管制, 使用者不能自由發言或發言即刻被刪除, 在研究社群網站資料時容易受阻 (Bamman et al., 2012)

以下表 2 與圖 5 為三大社群網站的功能比較, 在發文對象的功能來看, Facebook 與 Instagram 都有限制能看到該篇文章的對象, 而 Twitter 則是公開的。也因此, 在情報傳播速度這方面, Twitter 遠快於其他兩者, 並且能快速獲得更多新粉絲。其他兩者的粉絲鞏固舊粉絲為主。

表 2:三大社群網站比較表

	Facebook 	Twitter 	Instagram 
全球用戶	20 億	3.28 億	8 億
按讚功能	有	有	有
分享功能	分享	轉推 (Retweet)	無
發文對象	朋友、朋友的朋友	公開	追隨者
文章內容	文字+圖片影片	文字+圖片影片 (字數限制 140 字)	圖片、影片為主
訊息擴散性	中	高	低
粉絲互動	強化與現有的粉絲關係	獲得新粉絲	強化與現有的粉絲關係

資料來源: Social Media Lab

情報擴散速度

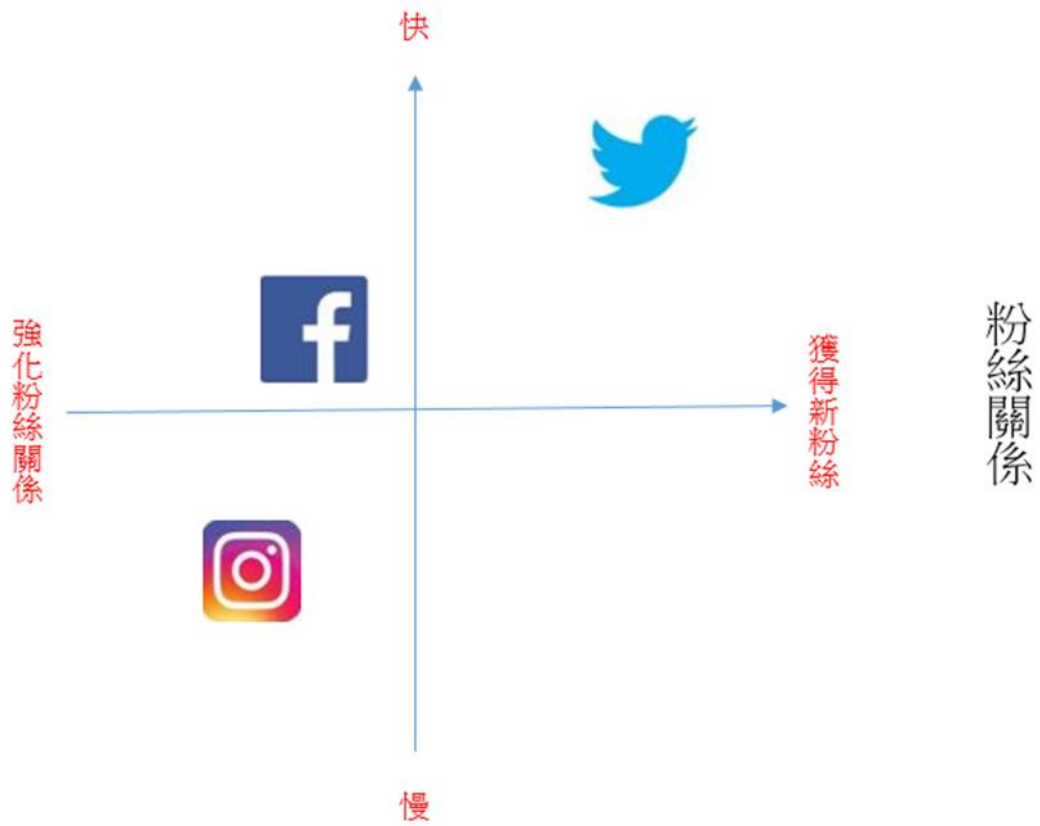


圖 5:三大社群網站情報擴散速度比較

資料來源:Social Media Lab,<https://gaiax-socialmedialab.jp/post-29375/>

第三章 文獻探討

本章節共分為三個部分，第一節將介紹幾個較知名的社群網站以及過去以社群網站為對象的研究文獻，包含粉絲專頁的經營數據，或是以發文、按讚數等數據來做分析的文獻，經過歸納整理之後並做成表格。第二節為本研究的理論基礎，其中包含 AIDA 與 HoE 兩大模型，並基礎介紹這兩個模型的理論以及曾經做過的相關研究，理論與社群網站發文的關聯性等等，作為研究分析之參考。最後第三節將會介紹並解釋大數據的概念以及大數據的應用。

第一節 社群網站相關研究

社群網站的主要作用是為一群擁有相同興趣與活動的人建立線上社群。這類服務往往是基於網際網路，為用戶提供各種聯繫、交流的互動通路，如電子郵件、即時通訊服務等。此類網站通常通過朋友，一傳十、十傳百、百傳千地把網路展延開去，就像樹葉的脈絡，華語地區一般稱之為「社群網站」(Wikipedia)。社交媒體的社群組成與網路論壇的社群組成亦有差異，由小眾社群變成大眾參與。以往的網路社群成員有明顯的共同興趣或討論議題為其認同基礎，群體規模從數十到數百人，共同在特定的虛擬空間進行互動。但社交媒體串連起的是沒有固定邊際的社會網絡群體 (networked community)，創造了一個多元、流動的空間，讓使用者可在龐大的社交網絡任意與人群接觸，連結眾多異質性的網絡社群 (翟本瑞，2011)。

Facebook 是目前全球最大的社群網站。由 Mark Zuckerberg 於 2004 年 2 月 4 日創立，希望透過提供使用者分享平台的方式，達到讓世界更加開放與廉潔的使命(Facebook, 2004)。之後在 2007 年推出 Facebook 粉絲專頁，企業與品牌或組織等團體可以透過粉絲專業分享動態與情報，進而達到與使用者對話的目的

(Facebook, 2016)。

根據 techcrunch 在 2017 年做的調查，Facebook 的月活躍用戶已經突破 20 億。若將 Facebook 視為一個國家，其月活躍人數已成為全球最大國，超越中國的 13.8 億人口，與印度的 13.2 億人口 (GEOHIVE, 2016)。在台灣，有超過 75% 的人使用 Facebook；以台灣連網人口計算，則高達 9 成以上，以 Facebook 台灣用戶計算：1300 萬每日活躍用戶數中，有 1200 萬都用手機連 Facebook，滲透率達 92% (蕭文康, 2015)。根據皮尤研究中心 2015 年初發布的「2014 年社群媒體更新報告」 (Social Media Update, 2014)，有約 7 成的美國青少年也在使用 Facebook。

Instagram 由 Kevin Systrom 與 Mike Krieger 建立，並於 2010 年 10 月首次推出。該服務迅速得到普及，截止 2018 年元月為止，註冊人數已經突破 8 億。用戶可以在 Instagram 上分享相片及視訊，也可關注其他用戶的分享。用戶在將自己的帳號與其他社群網路帳號連結後亦可分享至其他社群網站 (Wikipedia)。根據資訊工業策進會創新應用服務研究所 (FIND) 統計，Instagram 在台灣的使用者比例約 3 成 (32.7%)，其中以 12 歲到 24 歲的青少年使用率遠超過其他年齡層，顯示 Instagram 在青少年當中的人氣 (FIND, 2015)。

Twitter 是一個社群網路與微網誌服務，它可以讓用戶更新不超過 140 個字元的訊息，這些訊息也被稱作「推文 (Tweet)」。推特成立於 2006 年，截至 2017 年，Twitter 共有 3.28 億月活躍用戶。Twitter 被形容為「網際網路的簡訊服務」。網站的非註冊用戶可以閱讀公開的推文，而註冊用戶則可以通過 Twitter 網站、簡訊或者各種各樣的應用軟體來發布訊息 (Wikipedia)。根據 FIND (圖 6) 的統計調查，Twitter 在台灣的使用者比例相當低，僅約 3.8%，擁有 Twitter 帳號的比例也只有 16.3% (FIND, 2015)。國人擁有社群網站帳號比例最高的是 FB 與 LINE，

LINE 是韓國開發的即時通訊軟體，目前已經有超過 10 億人註冊使用，在全球通訊軟體用戶數排行中名列第七。由於智慧型手機的進步，LINE 不只能傳送文字與圖片，還能發送影片、動畫、語音通話等功能。

根據 Pew Research Center 於 2015 年 7 月發布的最新報告，美國青少年有約 3 成使用 Twitter。另外有 63% 的美國人使用 Facebook 與 Twitter 收看新聞，特別當突發事件發生時，有 59% 人會用 Twitter、31% 的人用 Facebook 追蹤事件的最新發展，主因是 Twitter 在即時訊息的傳遞比 Facebook 更快，以時序而非演算法來篩選訊息是 Twitter 與其他社群網站較為不同的特色 (Pew Research Center, 2015)。

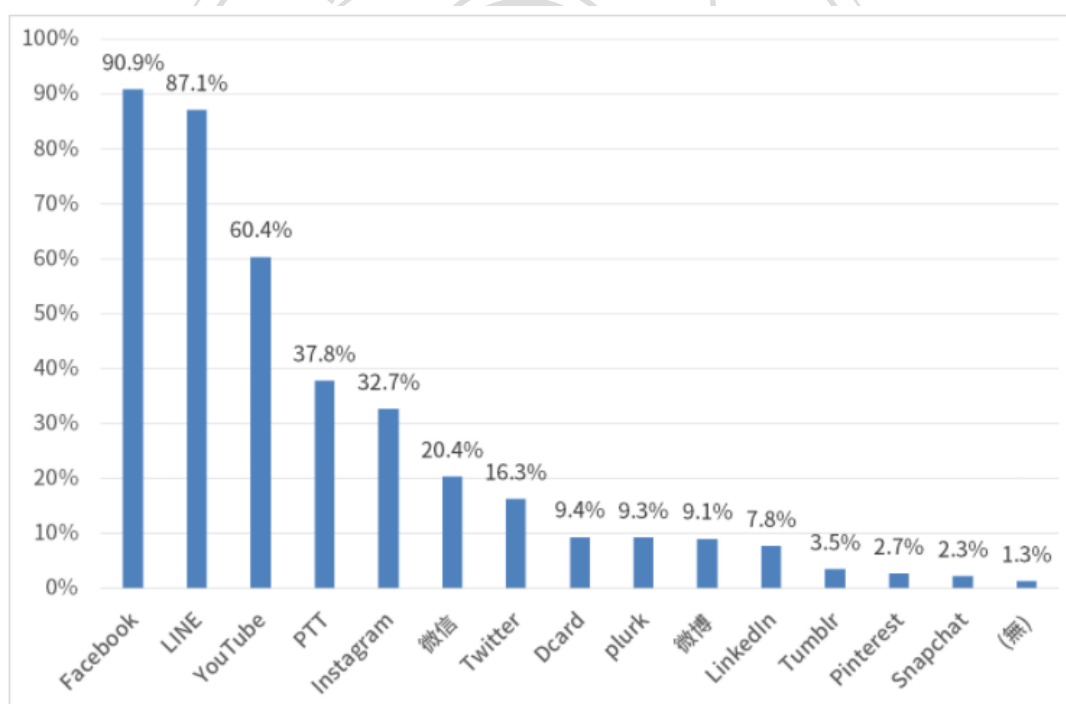


圖 6: 國人擁有社群帳號比例

資料來源：資策會 FIND (2016H2)

以社群網站為研究對象的國內外的已經有不少了，但其中涉及粉絲專頁經營操作的研究就比較少（余峰璋，2017）。張哲豪（2014）的研究就以 Facebook 的粉絲專業經營為研究對象，以 Zhao & Morad (2013) 所提出社群商務的五個面向（個人、交談、社區、商務，與趣味性）與三個指標（關注度、參與度，與傳播度）分析高績效的粉絲專頁該有哪些要素。結論是：1.內容相符、焦點話題、連結數、發文數、鼓勵回覆，以及邀請使用其他應用程式等要素對粉絲互動影響較顯著。2.鼓勵粉絲分享、轉傳消息所得到的粉絲互動效果不佳。

Huber et al. (2011) 利用 Facebook 粉絲專頁上公開與非公開的數據分析：哪些方面而努力使企業經營的粉絲專頁擁有更多的活躍粉絲。研究時間為 2010 年 10 月至 2011 年 9 月，共一年。變數包含：總粉絲數、平均每日發文數、平均每日回文數、含有連結、圖片，或 app 的文章數。研究結果顯示：1.總粉絲數越多、管理者每日貼文數越多，以及管理者與粉絲的互動越多都能增加粉絲專頁的活躍粉絲數。2.含有圖片或 app 的發文會比含有超連結的發文吸引更多活躍粉絲。

林質安和劉蕙苓（2016）則以修改的 Kent Taylor 「對話式溝通」(dialogic communication)分析臺灣現代戲劇團隊經營 Facebook 粉絲專頁時所運用的公關策略，以及各策略所獲得的回應情形。研究結果顯示，全部劇團於資訊實用性策略上的表現極佳，鼓勵參與次之，留住訪客較低。而大型劇團較擅長於資訊管理宣傳部分；小劇團則注重與粉絲的互動。林質安、劉蕙苓認為，要達到雙向平等對話溝通的理想，粉絲人數與公關資源必須調節至「最佳雙向互動規模」便是最佳狀態。

林永鑫等（2016）以大數據分析台灣幾位演藝圈名人經營 Facebook 粉絲專頁的情形發現，粉絲對於藝人的私生活感情狀態的文章關注度最高，而與粉絲互動的文章則能獲得最多的留言數。

除了以粉絲專業經營為研究對象之外，社群網站中較常被拿出來研究的資料還有使用者的發文。陳香伶 (2014) 運用內容分析法與深度訪談法分析了包含 ETtoday 新聞雲、Yahoo!奇摩新聞、中天快點 TV、蘋果日報與自由時報的新聞粉絲專頁，探討假油事件的報導型態、主題框架與呈現方式，最後發現不斷重複的新聞議題會導致用戶對於議題的無感。

吳文貴和林孟陞 (2013) 以 Facebook 之粉絲專頁訊息內容為研究對象，探討 FB 粉絲專頁之廣告文字訊息，包括訊息標題類型、內容訴求類型對溝通效果之影響。最後選定「7-11 粉絲專頁」、「麥當勞粉絲專頁」為研究標的。結果顯示，文章標題類型、內容訴求與內容涉入程度對於廣告效果有不同程度的影響，但無交互作用；其中幾個重要影響：祈使式標題與感嘆式標語廣告效果較佳，內文訴求則以功利訴求的廣告效果較佳。而內文的涉入程度越高，廣告效果也會越好。

在各國總統大選期間，各大社群網站也會湧入大量討論關於選舉的話題，因此這段時間內的社群網站使用者的活動也是常常被拿來做研究的主題。Burgess & Bruns (2012) 以 2010 年澳洲大選為主題，探討選舉期間使用者如何在 Twitter 進行討論。他們蒐集了選前一個月到選後 2 天 (38 天) 所有標示 #ausvote 的推文，共有 41.5 萬則，結果發現有 22% 的 tweets 集中在選舉當天發出，在所蒐集到的共 3.7 萬發文者中，有 1.9 萬人 (51%) 只在選舉當天發文，也就是有一半的參與者只在開票結果出爐當天才對此事發表意見。他們進一步分析發現，這些參與選舉議題討論的活躍使用者反映了特殊次文化，通常是網路重度使用者對科技政策、新聞和資訊深感興趣而常積極參與討論，以打趣嘲諷的語氣評論主流報導或政治人物，並非真的投入嚴肅議題討論。

Groshek & Al-Rawi (2013) 研究 2012 年美國總統大選期間，針對候選人 Obama 與 Romney 的 Facebook 粉絲專頁與含有 #election2012 標示的超過 100 萬

條 twitter 推文進行文字探勘，並排名出兩陣營與民眾討論中最常出現的幾個詞彙，最後以概念圖 (Concept map) 的方式呈現。蒐集資料的時間從 2012 年 1 月 1 日到 2012 年 11 月 7 日 (選後一天)。

鄭宇君和陳百齡 (2014) 研究台灣 2012 總統大選在中文 Twitter 圈的討論資料，探討這些不同語言社群對台灣總統大選的關注程度。資料蒐集主要透過關鍵字搜尋，其中包含各候選人名字、政黨等關鍵字，最後以視覺化工具繪出社交網路關係圖並分析各國網友對同一個議題討論的緊密程度。該研究另外將華人圈內幾個追隨數 (Followers) 較高的帳號列出做為總統大選中發言的代表性使用者。研究結論提到，Twitter 做為全球社交網站，資訊傳播模式深受在地脈絡影響，包括使用者與事件位置的鄰近程度、使用者與社群關係的緊密程度。

因社群網站有著快速傳遞訊息的特色，再加上近年來行動網路快速普及，社群網站也常被用來作為社會運動的工具，不論是號召民眾，或是替在場的參與者聲援打氣。例如 Papacharissi & Oliveira (2013) 以阿拉伯之春時期的三個月內，利比亞用戶的 Facebook 及 Twitter 發文做分析，其中包含三個粉絲專頁的發文。內容包含：捍衛人權、殘害人權、國際援助、社會動員等話題。其中這些話題呈現了四個特徵：即時性 (instantaneity)、參與群眾外包的菁英 (crowdsourced elites)、團結 (solidarity)、環境氛圍 (ambience)。結果顯示：由於公民記者和目擊者透過 Facebook 與 Twitter 將利比亞的資訊快速傳播，使得參與者迅速增加。Papacharissi & Oliveira (2012) 認為，這些社群網站有別於傳統新聞，帶給現場群眾一個更快速傳播消息的管道，並且能在發文中夾帶較情緒性的文字，因此能快速增加情感的累積。

Thorburn (2014) 探討社群網站直播對社會運動的重要性。他以 2012 年加拿大魁北克學生罷課運動為例，原先做為校園與社區影音頻道的 CUTV 在罷課運

動期間成為網路直播頻道，持續直播學生們的占領行動。它提供了街頭政治行動者一個建構主體性的位置。Thorburn 認為，線上直播功能在社運中產生了革命性影響，參與民眾和視聽者在多樣化的行動中，透過監看與被監看形成了反霸權的監看組合（counter-hegemonic surveillance assemblage）之可能性。

鄭宇君和陳百齡 (2016) 則是以台灣 2014 年 3 月 18 日發生的 318 佔領國會抗爭運動為主題，利用 Twitter 做資料收集，藉由社群媒體鉅量資料的收集與事後分析，探討 318 運動期間的 Twitter 各種語言相關推文 (tweets) 的趨勢變化。該研究也透過縮網址還原技術 (unshorten URL) 將大量推文中帶有的縮連結還原為原始網址，進行超連結分析。透過超連結分析，該研究發現大多數 Twitter 用戶在 318 學運期間轉發的資訊來自 Twitter 的其他推文或是其他社群網站，包括 Facebook、Plurk。另一類則來自影音粉享網站，如一般大眾常用的 YouTube、影音直播平台的 Ustream.tv 直播頻道或是日本 niconico 生放送 (live.nicovideo.jp)。

表 3: 社群網站研究文獻回顧

文章標題	作者	變數與研究結果
企業粉絲專頁之品質要素分析	張哲豪	以 Facebook 的粉絲專業經營為研究對象，分析高績效的粉絲專頁該有哪些要素。結論是:1.內容相符、焦點話題、連結數、發文數、鼓勵回覆，以及邀請使用其他應用程式等要素對粉絲互動影響較顯著。2.鼓勵粉絲分享、轉傳消息所得到的粉絲互動效果不佳。
USING FORUM AND SEARCH DATA FOR SALES PREDICTION	Hube, Landherr, Probst, Reisser	利用 Facebook 粉絲專頁上的數據分析:哪些方面的努力使企業經營的粉絲專頁擁有更多的活躍粉絲。

OF HIGH INVOLVEMENT PRODUCTS		1.總粉絲數越多、管理者每日貼文數越多，以及管理者與粉絲的互動越多都能增加粉絲專頁的活躍粉絲數。2.含有圖片或 app 的發文會吸引更多活躍粉絲。
宣傳為先？互動為主？ 台灣現代劇團經營 Facebook 粉絲專頁之公關溝通策略應用分析	林質安、劉蕙苓	利用來自粉絲的按讚數、留言、分享數以及來自其他劇團的按讚數、留言、分享數與該粉絲團經營的貼文數、發布圖片影片數與回覆數作為變數。分析大小劇團的粉絲劇團公關策略差異。
台灣名人經營粉絲專頁成功因素之研究-以 Facebook 為例	林永鑫、黃耀賢、林奎仟與楊濠鴻	該研究以知名藝人的粉絲專頁發文數與發文類型以及粉絲按讚數做分析。結果發現粉絲對於藝人的私生活感情狀態的文章關注度最高，而與粉絲互動的文章則能獲得最多的留言數。
新聞媒體 Facebook 粉絲專頁之守門、呈現與使用-以 2013 年臺灣食用油造假事件新聞為例	陳香伶	運用內容分析法 (包含報導類型、主題類型、報導呈現方式、轉貼新聞數及字數等變數)與深度訪談法分析新聞網站報導型態。最後發現不斷重複的新聞議題會導致用戶對於議題的無感。
Facebook 粉絲專頁訊息廣告之研究:以 7-Eleven 與麥當勞粉絲專頁為例	吳文貴、林孟陞	以粉絲專頁發文的標題類型、內容訴求類型、訊息涉入程度以及廣告效果為變數進行分析。祈使式標題與感嘆式標語廣告效果較佳，內文訴求則以功利訴求的廣告效果較佳。而內文的涉入程度越高，廣告效果也會越好。
(Not) the Twitter election: The dynamics of the #ausvotes conversation in relation to the Australian media ecology.	Burgess & Bruns	以 2010 年澳洲大選為主題，探討選舉期間使用者如何在 Twitter 進行討論。該研究分析 hashtag #ausvotes 在大選期間各項數據的變化如文章數、用戶數。分析發現，這些參與選舉議題討論的用戶通

		常對科技政策、新聞和資訊感興趣而積極參與討論。以幽默嘲諷的語氣評論主流報導或政治人物，並非真的投入嚴肅議題討論。
Public Sentiment and Cultural Framing in Social Media Content During the 2012 U.S. President Campaign	Groshek, J. & Al-Rawi, A.	研究 2012 年美國總統大選期間，針對候選人 Obama 與 Romney 的 Facebook 粉絲專頁與含有 #election2012 標示的推文分析，主要以字詞分析排列出兩陣營與其他用戶最常使用的幾個詞彙並進行關聯分析。
探索 2012 年台灣總統大選之社交媒體浮現社群：鉅量資料分析取徑	鄭宇君、陳百齡	研究台灣 2012 總統大選在 Twitter 的討論資料，包含各語言的推文數與討論人數變化以及知名人物的發文數。研究結論提到，Twitter 做為全球社交網站，資訊傳播模式深受在地脈絡影響，包括使用者與事件位置的鄰近程度、使用者與社群關係的緊密程度。
Affective news and networked publics: The rhythms of news storytelling on #Egypt.	Papacharissi, Z., & de Fatima Oliveira, M.	該研究收集了包含 #egypt、#Jan25、#Tahrir 等 hashtag 推文進行篇章分析。研究認為，這些社群網站有別於傳統新聞，帶給現場群眾一個更快速傳播消息的管道，並且能在發文中夾帶較情緒性的文字，因此能快速增加情感的累積。
探索線上公眾即時參與網絡化社運—以臺灣 318 運動為例	鄭宇君、陳百齡	以 318 學運為主題分析該時段內的推文趨勢與附帶的超連結來源是源自哪些網站。變數包括推文聲量變化、超連結分析以及台灣新聞媒體被引用情形。結果發現，該時段在推特上轉傳的資訊通常都是來自其他社群網站。

資料來源:本研究整理

第二節 AIDA 模式

愛達 (AIDA) 模式是 Elmo Lewis 在 1898 年所提出描述消費者行為的模式，並在 1925 年由 Strong, E.K. 將 AIDA 模式導入廣告評價，成為第一個廣告效果測量模式 (曾碧美, 2014)。AIDA 代表著消費者行為四個步驟的縮寫，其四個步驟依序是：關注 (Attention)、興趣 (Interest)、慾望 (Desire)、行動 (Action)，多用於行銷相關領域以及廣告效果之衡量。

Kono (2009) 認為隨著網路科技快速發展，因科技的發展在消費者行為上有三項重大的改變如下：1. 搜尋行為的變化：消費者能夠透過網路，迅速獲取自己需要的資訊，進而找到商品，在資訊的獲取上比以前要在容易許多。2. 訊息可由消費者自行發出：消費者可以在社交網站上對商品進行評論，而這些評論將影響其他消費者的購買決策。3. 購買行為的變化：消費者接收訊息之後，若消費者喜歡可以立即透過電子商務網站進行購買，若消費者打算到實體店面購買，也可利用網路找到距離最近的商店。

因此，基於以上三要素，Kono 將 AIDA 模式改良並提出 AISAS 新消費者行為模式，其中五個字母代表：關注 (Attention)、興趣 (Interest)、搜尋 (Search)、購買 (Action) 與分享資訊 (Share)。而趙滿玲 (2014) 則將 AIDA 與 AISAS 結合，提出 AISDAS 的概念，其中的六個縮寫為：「提起」其他消費者注意 (Attention)、興趣 (Interest)、搜尋 (Search)、慾望 (Desire)、購買 (Action)、分享資訊 (Share)。而分享資訊又會引起其他人注意，因此最後一步驟又會連接到第一步驟，形成一個循環，如圖 7。

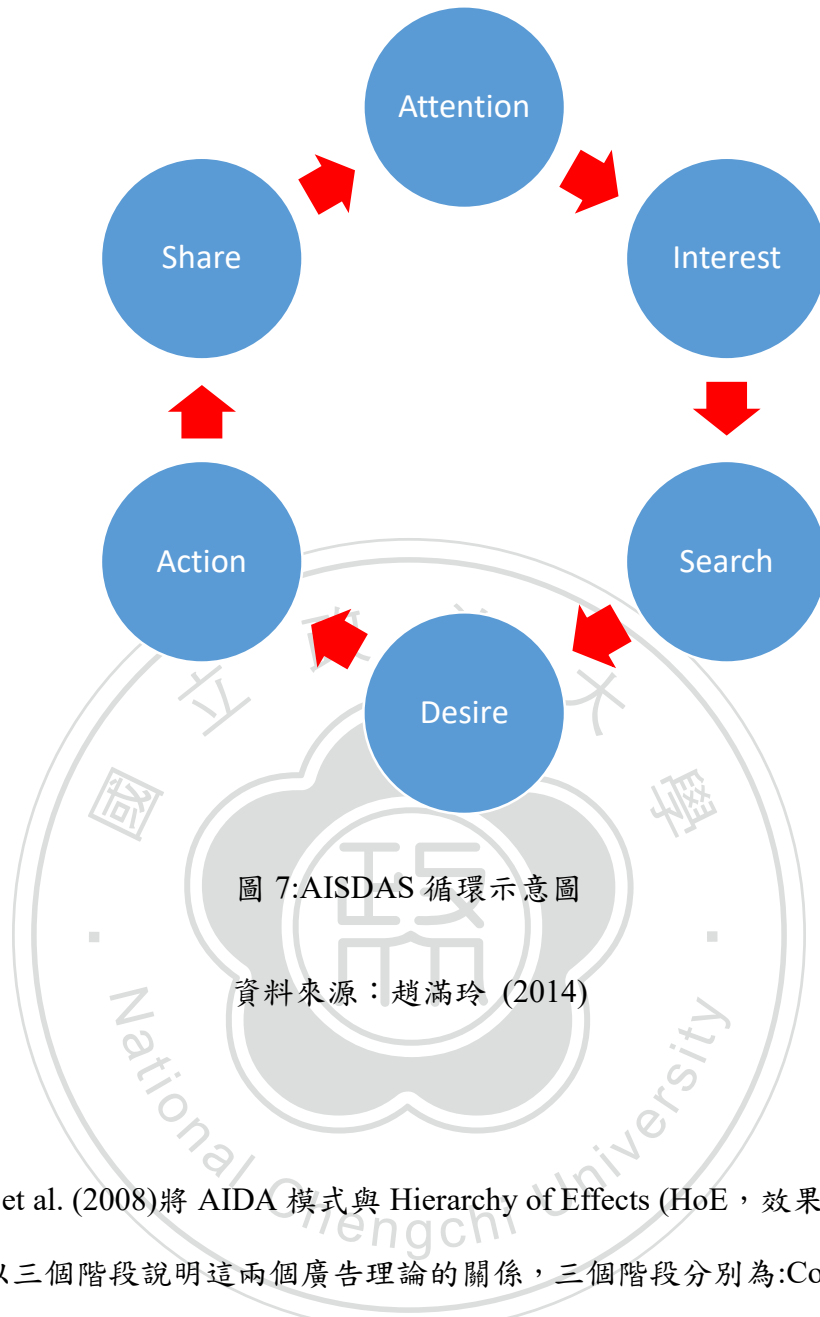


圖 7:AISDAS 循環示意圖

資料來源：趙滿玲 (2014)

Belch et al. (2008)將 AIDA 模式與 Hierarchy of Effects (HoE，效果層級模式) 結合，並以三個階段說明這兩個廣告理論的關係，三個階段分別為:Cognition (認知)、Affect (情感)，以及 Behavior (行為)。HoE 最早是由 Lavidge 和 Steiner 建立 (Lassen et al., 2014)，並可分為六個階段，分別是 Awareness (知曉)、Knowledge (了解)、Liking (喜愛)、Preference (偏好)、Conviction (確信)，以及 Purchase (購買)，這六個小階段分別又對應到認知、情感，及行為三個大階段 (蕭富峰等，2010)。而 AIDA 模式也能被整合到這三個階段內，如下表 4 所示。

表 4:AIDA 與 HoE 模式比較表

	AIDA	HOE
認知	Attention	Awareness Knowledge
情感	Interest Desire	Liking Preference Conviction
行為	Action	Purchase

資料來源: Lassen, Madsen, Vatrapu (2014)

Lassen et al. (2014) 認為在社群網站上的使用者對某個商品感興趣而發表的文章可以代表該使用者處於 AIDA 模式或 HoE 模式中的某個階段。例如:我今天注意到朋友換了新手機 (Attention、Awareness 階段), 我比較了幾支新手機的性能 (Interest、Knowledge、Liking 階段)、我買了新手機 (Action、Conviction、Purchase 階段) 等等。

洪培修 (2013) 以 AIDA 模式的四階段為研究主題, 在網路廣告效果的研究中發現在 Yahoo 橫幅廣告上廣告的追蹤性和目標客群接觸性有助於引發消費者注意和產生興趣的效果, 互動溝通性則會有提升消費者對購物慾望與行為的效果。在 Facebook 社群網站廣告方面, 傳遞的彈性則有助於提升廣告閱聽人引發注意和產生興趣的程度, 以及在社交共享性中上有助於提升購物慾望和行動的效果。研究結果發現在引發注意與品牌知名度、產生興趣與品牌聯想、引發慾望與品牌知覺品質以及付出購買行動與品牌忠誠度有著正面的影響。因此運用不同類型的廣告在 AIDA 模式當中的各階段, 對於品牌權益當中不同的構面上, 能提升廣告產生的效果。

第三節 大數據相關研究

巨量資料 (Big data)，又稱為大數據，指的是傳統資料處理軟體不足以處理它們的龐大或複雜資料集的用語，這些資料集大小常超出人類在可接受時間下的收集、度用、管理和處理能力。在總資料量相同的情形下，與個別分析獨立的小型資料集 (Data set) 相比，將各個小型資料集合併後進行分析可得出許多額外的資訊和資料集間的關聯性，可用來察覺商業趨勢、判定研究品質、避免疾病擴散、打擊犯罪或測定即時交通路況等；這樣的用途正是大型資料集盛行的原因 (維基百科)。嚴格來說，到目前為止，大數據的大小還沒有一個準確的定義，通常的說法是，「當資料量超過典型資料庫工具的硬體環境和軟體工具所能獲得、儲存、管理以及分析的能力」，就稱為「大數據」或是「巨量資料」(胡世忠，2013)。但是，許多研究者指出，大數據的真正關鍵其實不在資料量的大小，因為事實上許多資料集可能不大於 1 TB，而是研究方法與思維的改變，鉅量資料分析強調的是「分析全部資料」或「至少是盡量完整的資料」(包括其後設資料 metadata)，以便研究者查看資料中的細節或進行新的分析，發掘一些用其他方法分析無法得知的現象(城田真琴，2012／鐘慧真、梁世英譯，2013；Mayer-Schönberger & Cukier, 2013／林俊宏譯，2013)。Boyd & Crawford (2012) 也提到，巨量資料的出現不只代表採用了大量的數據、分析工具和研究流程而已，更重要的是重新形塑了知識建構與研究流程，還有人們應該如何處理並分類資訊。在思維與研究方面則產生了「向運算轉」(computational turn) 的趨勢，此影響由資訊、生物領域擴及人文社會研究領域。

關於大數據的定義，目前的定義普遍認為來自於 3V。3V 指的是容量 (volume)、速度 (velocity) 以及多樣性 (variety)，有別於傳統量化研究的資料，大數據分析的資料來源更為多元且數量更龐大，另外來源通常非結構性的且資料更

新速度很快 (Jordan & Lin, 2014; 李欣宜, 2015)。Hair & Mayer-Schönberger (2014) 也提到, 研究者與企業間以 3V 來形容大數據的特性之外, 同時隱含了高準確性與高可靠性的意思。3V 與後來提出的第四個 V, 也就是真實性 (Veracity) 合稱為 4V (Beyer & Lancy, 2012)。該名詞是由 Inderpal Bhandar (Express Scripts Chief Data officer) 在波士頓大數據創新高峰會 (Big Data innovation Summit) 的演講中提出。由於大數據資料過於雜亂, 分析之前須過濾資料偏差、偽造或異常的訊息, 避免「Dirty Data」干擾到資料系統的正确性, 進而影響決策結果 (國立台灣大學統計教學中心)。

社交媒體產生的巨量資料也對傳播領域的研究帶來了新的挑戰, 一方面是研究者如何取得傳輸資料 (transactional data), 因為這些數據大部份被 Facebook, Twitter, Google 等大公司掌握在手裡; 另一方面, 鉅量資料的搜集與分析需要資料專家的運算技術與人文社會學者的領域知識的整合, 因此必須透過跨領域合作發展適當的社交媒體分析工具, 縮減資料專家與未經資料訓練的人文研究者間的「資料分析落差」(data analysis divide), 才能從鉅量資料分析中找尋洞見 (Manovich, 2011; 鄭宇君和陳百齡, 2014)。

Asur & Huberman (2010) 利用 Twitter 蒐集到約 300 萬則推文資料 (如推文數、轉推數、超連結) 與 HSX (Hollywood Stock Exchange) 指標為研究對象, 試圖預測電影的票房收入。研究結果顯示, 他們搶先在電影上映前就成功地預測了在 2009 年到 2010 年間的 24 部電影的票房, 顯示社群網站可以被用來預測未來的結果。

Lassen et al. (2014) 則以 Asur 和 Huberman 的研究方法出發, 蒐集了 2010 年到 2014 年 Twitter 上有關 Iphone 的推文資料建立模型, 並與同時期的 Iphone 實際銷售數據做比較, 最後也成功以 Twitter 的資料預測了 Iphone 的下一期的銷售

量。該研究認為:tweets (推文) 可以作為社群網站使用者對某商品表達興趣或購買意圖的媒介。他們也認為這套預測的理論能夠應用於其他在社群網站上被熱烈討論的商品。

Zimbra et al. (2009)以跨國零售業沃爾瑪 (Walmart) 的相關討論版為研究對象，利用大數據與文本情感分析，分析討論區內的發文內容組成與情感分數。並將討論區內發文數最多的前幾名意見領袖 (opinion leader) 獨立做分析。研究發現，在討論區裡最活躍的前 1%使用者就回覆了高達 30%的文章量；前 10%則共回覆了討論區內 70%的文章數。

Wu & Brynjolfsson (2015)則是將 Google 搜尋的大數據加入到預測房價與銷售的模型裡。利用 Google 趨勢 (Google Trend) 提供的搜尋熱度數據能更精準地預測房價與銷售量。該研究結論裡提到，利用 Google 趨勢做預測不只能應用在房價上，也能運用在 DVD、Blu-ray，或其他電子產品上，甚至是勞動市場。

Nann et al. (2013) 則利用部落格、社群網站以及討論版的數據與文本情感分析分數來分析與預測 Standard & Poor's 500 index (S&P 500) 的每日變動，研究期間為 2011 年 6 月至 2011 年 11 月，共半年。結果也顯示，民眾在網路上討論的數據能夠預測股價變動，也驗證了這些網路討論的聲量在金融市場的分析是有價值的。

Tzur et al. (2014) 蒐集社群網站上的推文進行分析，並分析使用者在社群網站上對交通政策的意見與最頻繁討論的方面等。該研究利用文字探勘技術將推文中最常提到的詞彙排列出來並分類，研究結果顯示，使用者在社群網站上最關注的是交通運輸的品質，特別是公車和火車。交通運輸的價格則是較少人關注的議題。

劉嘉薇 (2017) 蒐集了包含社群網站、討論區，以及部落格等超過 7000 個網站頻道，其中包含 Facebook、Youtube、PTT，與 Mobile01 等國內大型網站的資料，進行網路對於台灣統獨立場與聲量大小的研究。研究中發現，不論是支持統一或是獨立的話題，正面情感分數與負面情感分數的比例都低於 1，顯示在討論這類話題時，用戶的情緒都較為負面，例如包含謾罵、攻擊另一族群的言論等。另外可以發現，支持統一或支持獨立雙方的意見領袖活躍的論壇也不一樣，例如支持統一的一方在卡提諾王國與伊莉討論區較活躍；支持獨立的一方在 PTT 八卦版與 Facebook 上較活躍。

謝元晟等 (2016) 利用 R 軟體建立文字探勘平台，並蒐集電視節目《嫁妝》的 Facebook 粉絲專頁的相關數據 (如:按讚數、分享數、留言數、粉絲團人數與文字探勘之分析結果所得到的三個變數，包括情感分數、正面情感語句、負面情感語句等七個變數)預測收視率與其他關聯分析。資料蒐集時間為 2015 年 4 月 1 日至 2015 年 5 月 29 日，共 60 天。該研究模型的模型預測能力檢驗參數 MAPE 為 6.222 小於 10，顯示該模型能精準預測《嫁妝》這個節目的收視率。利用社群網站的資料進行預測還能以小時或分鐘為區間，分析出及時的收視率變化。

表 5: 大數據研究文獻回顧

文章標題	作者	變數與研究結果
Predicting the Future With Social Media	Asur & Huberman	利用 Twitter 蒐集到約 300 萬則推文資料與 HSX 預測 24 部未上映的電影票房。變數包含推文數、上映戲院數、PNratio、附帶超連結比例及 HSX
Predicting iPhone Sales from iPhone Tweets	Lassen, Madsen, Vatrapu	收集 Twitter 上有關 Iphone 的推文資料建立模型，並與同時期的 Iphone 實際銷售數據做比較最後預測下一期 Iphone 的銷量。變數包含推文數、PNratio
Assessing Public Opinions Through Web 2.0: A Case Study on Wal-Mart	Zimbra, Fu, Li	以沃爾瑪的相關討論版上的文章進行分析，透過文章內容與文本情感分析歸類出討論文章的類型。並將幾名具代表性的意見領袖的文章獨立分析，並比較其與一般使用者的不同。變數有情感分析模型與意見領袖 (Opinion Leader) 的推文數、回文數以及情感分數等。
The Future of Prediction How Google Searches Foreshadow Housing Prices and Sales	Wu & Brynjolfsson	該研究將 Google 搜尋的大數據加入到預測房價與銷售的模型裡。利用 Google 趨勢提供的搜尋熱度數據能更精準地預測房價與銷售量。
Predictive analytics on public data-The case of stock markets	Stefan, Jonas, Detlef	蒐集來自部落格、社群網站以及討論版的數據與文本情感分析分數建立模型進行分析與預測 Standard & Poor' s 500 index (S&P 500) 的每日變動。
The potential of social media in delivering transport policy goals	Tzur, Muller, Kuflik, Minkov, Nocera, Shoor	蒐集社群網站上的推文進行分析，並分析使用者在社群網站上對交通政策的意見與最頻繁討論的議題等。變數包括關鍵字與推文種類比例。該研究利用文字探勘技術將推文中最常提到的詞彙排列出來並分類。研究結果顯示，使用者在社群網站上最關注

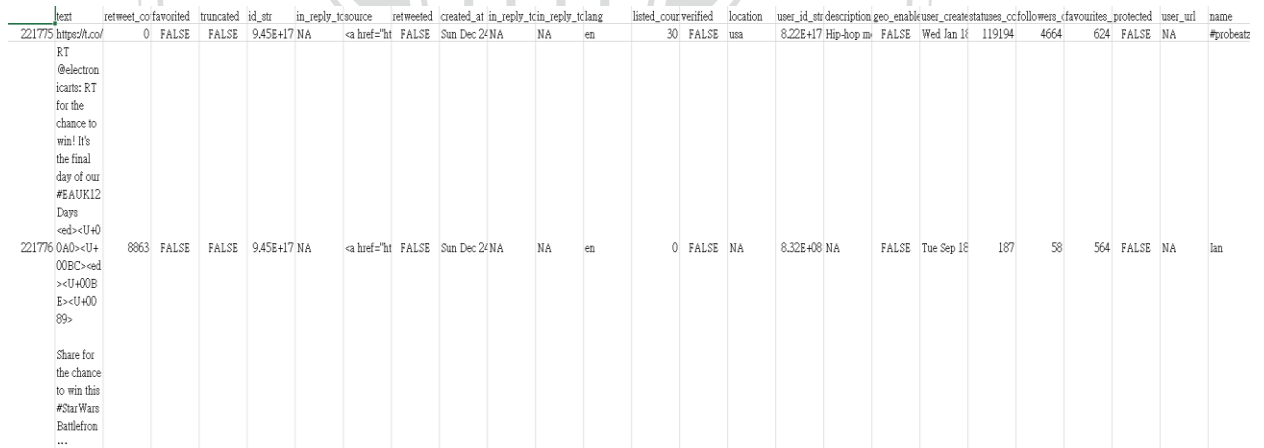
		的是交通運輸的品質，特別是公車和火車。交通運輸的價格則是較少人關注的議題。
網路統獨的聲量研究: 大數據的分析	劉嘉薇	蒐集了國內各種社群網站、討論區及部落格的文章進行網路統獨聲量分析。變數有統獨聲量及立場變化、常用詞彙類型以及正負面情緒 P/N 比。研究結果發現統獨雙方活躍的討論版也不盡相同。
運用 R 建立文字探勘平台應用於電視收視率預測	謝元晟、程美華、張光昭	利用 R 軟體建立文字探勘平台，蒐集 Facebook 粉絲團的各項數據來分析預測電視節目《嫁妝》的收視率。該研究使用集群分析、脈絡分析與情緒分析等方式分析推文內容並以官方粉絲團之按讚數、分享數、留言數、粉絲數與情感分數、正面情感語句、負面情感語句等變數建立預測收視率模型。

資料來源:本研究整理

第四章 資料說明

第一節 研究對象

本研究以英語區的 twitter 用戶作為研究對象，利用 R 軟體結合 twitter API 來抓取 twitter 上的推文。由於 twitter API 限制的關係，使用者只能夠抓到一星期以內的推文，因此本研究以每星期抓取一次的方式，自 2017 年 10 月 1 日開始每周蒐集一次資料至 2018 年 6 月 2 日結束，共約八個月的資料進行分析。蒐集的方式是以關鍵字的方式搜尋，R 軟體會自動將包含關鍵字 PS4 的推文蒐集起來並轉換成 csv 檔儲存，接著再以這些 csv 檔進行相關的統計分析。這些 csv 檔內包含推文內容、推文時間、帳號名稱、轉推數、like 數、該帳號的追隨者人數等等，在之後的研究結果中會將資料整理之後取出需要的部分進一步進行統計分析與視覺化繪圖。下圖 8 為剛收集到尚未經過整理的資料內容。



text	retweet_count	favorited	truncated	id_str	in_reply_to_source	retweeted	created_at	in_reply_to_in_reply_to_lang	listed_count	verified	location	user_id_str	description	geo_enabled	user_created_at	statuses_count	followers_count	favorites_count	protected	user_url	name		
RT @electron icants: RT for the chance to win! It's the final day of our #EAUK12 Days <ed><U+00BC><ed>	0	FALSE	FALSE	9.45E+17	NA	<a href="ht	FALSE	Sun Dec 24	NA	en	30	FALSE	usa	8.22E+17	Hip-hop m	FALSE	Wed Jan 11	119194	4664	624	FALSE	NA	#probeatz
221776 0A0><U+00BC><ed>><U+00BE><U+00E><U+00E9> Share for the chance to win this #StarWars Battlefront ...	8863	FALSE	FALSE	9.45E+17	NA	<a href="ht	FALSE	Sun Dec 24	NA	en	0	FALSE	NA	8.32E+08	NA	FALSE	Tue Sep 16	187	58	564	FALSE	NA	ian

圖 8:Twitter 推文屬性

資料來源:本研究整理

第二節 資料收集

R 語言在安裝了相關的套件 (如 streamR) 之後便可透過 twitter API 的設定開始抓取資料。在 twitter API 中將自己的帳號與專屬的 key 輸入到 R 軟體後便可以開始運作。而想抓取的資料的選擇也十分多元，例如可以設定讓程式收集多長時間的資料或是收集的一定數目推文之後便終止，也可以選擇專門收集哪一種語系的推文，例如:英文、中文、日文等。本研究在抓取推文時對收集資料的時間與推文數量都沒有加以設定，僅設定針對英文推文進行蒐集。圖 9 為 twitter API 的主介面。

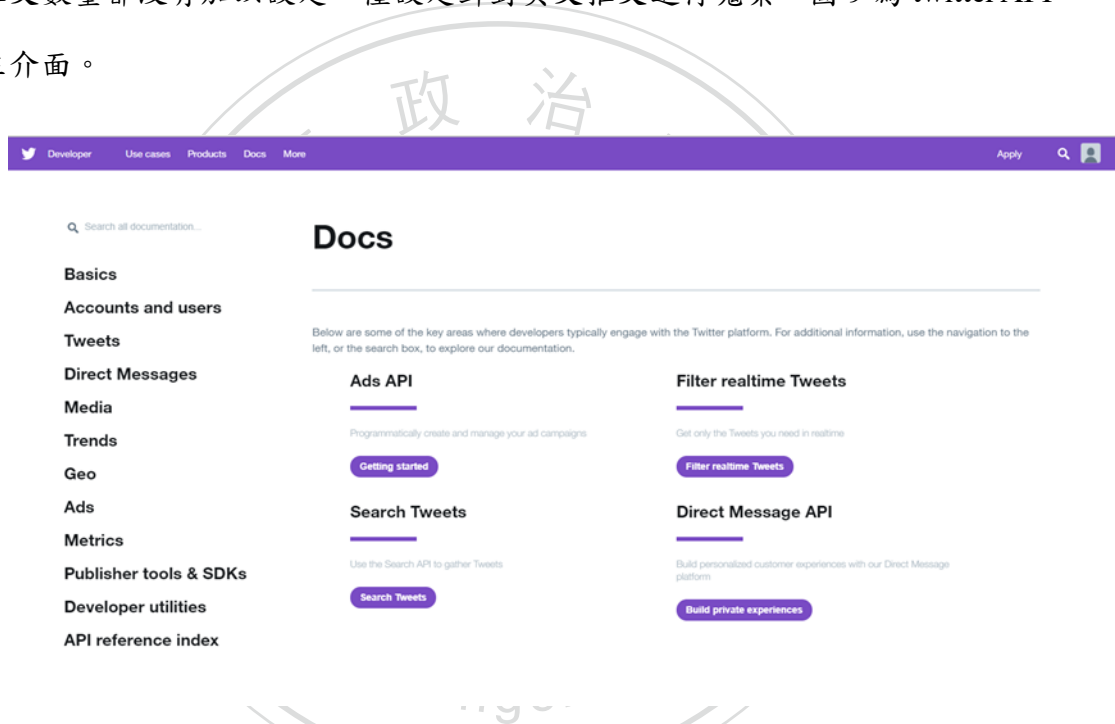


圖 9:twitter API 主介面

資料來源:Twitter 取自 :<https://developer.twitter.com/en/docs.html>

目前的社交媒體大數據研究以 twitter 佔大多數，因 twitter 的推文較公開且流動比其他社群網佔快，資料容易撈取而受到喜愛。Twitter 的資料撈取方式可分

為以下四種 (表 6):

一、以 hashtag 為撈取對象:使用者在推文時有時會在內文中加入與文章相關的 hashtag。如討論 ps4 相關議題時可蒐集包含#ps4 的所有推文;討論 xbox 的相關議題可蒐集包含#xbox 的所有推文,以此類推。

二、以關鍵字為撈取對象:透過在 R 軟體上設定要撈取的關鍵字, twitter API 便會每隔一定時間 (10-20 分鐘) 撈取一次包含該關鍵字的所有推文。之後透過地區、語言等方式篩選出研究者想要研究的特定對象。本研究使用的方式即為關鍵字撈取。

三、對特定用戶進行撈取:透過 twitter API 上設定特定用戶的帳戶 @ 帳號, 便可以針對該用戶撈取一段時間內與他相關的所有推文。由此研究者可設定一至多個使用者帳號, 再透過其朋友關係找出一連串相關的使用者帳號, 搜集這群帳號在特定時間內的推文 (鄭宇君和陳百齡, 2014)。例如, 網路社群研究領域的著名社會學者 Barry Wellman 以自己的 twitter 帳號搜尋, 找出和他自己有關聯的使用者, 他們同時與 Wellman 的相互追隨者當中至少有一個人以上的重疊。研究發現, twitter 中的社群既是真實的成員間彼此的互動, 又為想像的網路成員擁有社群意識, 且為社群提供資訊等 (Gruzd et al., 2011)。

四、撈取一定時間內所有推文: 由於這種方式必須採取超大樣本的資料分析方式, 資料量龐大, 主要是資科學者進行, 目的在於透過演算法建立社交媒體的傳播模型, 因此也經常有 Facebook, Twitter, Yahoo 的資料專家參與研究。如 Kwak et al. (2010) 透過資料演算方式分析使用者的傳播行為, 發現使用者將 Twitter 視為一種新聞媒介 (訂閱新消息) 遠大於將它視為社交網絡工具 (人際互動)。另外, Bond et al. (2012) 的研究則有 Facebook 資料專家參與, 在 2010 年以 6,100 萬

個 Facebook 用戶（約為當時 Facebook 用戶總數的十分之一）進行實驗，以瞭解線上政治動員與社會影響之效果，結果發現線上政治動員有效；但這麼大規模的資料搜集與分析誠非一般研究團隊所能負擔得起（鄭宇君和陳百齡，2014）。本研究選擇以關鍵字的方式來蒐集與 ps4 相關的推文，因研究對象為一個產品，因此以關鍵字搜尋能夠盡可能取得最多推文進行分析。

表 6: 蒐集資料方式比較表

方法	優點	缺點
Hashtag 撈取: 搜尋所有包含特定 hashtag 的推文並撈取。	容易蒐集到相關討論串。	有些使用者沒有用 hashtag 的習慣，無法蒐集到這些討論。
關鍵字: 在 R 軟體中設定特定的一個或一些關鍵字，撈取所有包含關鍵字的推文。	涵蓋的討論範圍最廣，資料數量最多。	需選擇合適關鍵字，否則可能蒐集到許多與研究無關的資料。
特定使用者: 針對特定使用者，搜尋所有與他相關的推文。	針對某些用戶研究他在社群網站上互動情形。資料較少。	資料數量較其他幾種方法少，只能分析少數人。
一定時間內所有推文	能夠蒐集到極大量的數據，並同時研究大多數使用者的行為。	成本過高，一般研究者無法負擔。

資料來源: 鄭宇君和陳百齡 (2014)

第三節 資料概述

本研究在 2017 年 10 月 1 日至 2018 年 6 月 2 日中，蒐集了包含 ps4 關鍵字的 523972 筆推文以及期間內所有 ps4 官方推特 (@playstation) 的文章進行分析，

如下表 7 所示。其中 PNratio 比每週平均為 2.83，推文中包含 url 連結的比例每週平均為 0.485。

表 7:本研究資料蒐集概況

	蒐集時間	發文數	PNratio	url 比例
Twitter 推文	34 週	523972 筆	2.83	0.485

資料來源:本研究整理

第四節 變數說明

本研究所選擇的被解釋變數為北美區 ps4 的周銷售量，資料來源為知名遊戲數據統計網站 VGChartz (<http://www.vgchartz.com/>)。VGChartz 會在網站上定期公布各家家用主機的交售量與遊戲銷售量，雖然更新速度稍慢，但在硬體銷售量方面的數據就較為準確，其中又以北美區的銷售數據最為完整。蒐集時間為 2017 年 10 月 1 日至 2018 年 6 月 2 日的數據。

解釋變數在一般使用者方面則包含發文數、正面文章比例 (PNratio)，AIDA 模型中認知、影響、表現這三期的每周文章比例，分類標準是根據關鍵字決定，詳見表 8。最後則是 google trend 的搜尋熱度。而在 ps4 官方專頁方面則包含 url 比例與官方與粉絲的互動率。發文數是將每周利用 R 軟體抓取的 twitter 文章經過整理後刪除與本研究不相關的文章後所得的實質文章數。PNratio 則是將整理過的推文再經過斷詞與刪除不必要的文字或標點符號後，進行文字情感分析並評分。再依據評分為正、負，或 0，將各篇文章分類為正面情感、負面情感，或中

性文章。而 PNratio 就是正面情感文章數除以負面情感文章數所得的比例。認知 (cognition)、影響 (affect)、表現 (behavior) 則分別代表每期文章有多少比例屬於這三期，分類方式為使用關鍵字分類。Trend 則是來自 google trend (<https://trends.google.com.tw/trends/>) 的搜尋熱度指數，指數越高代表在這段時間裡有越多人搜尋了這個關鍵字。使用 google trend 時地區設定只納入北美區的熱度數據，與本研究其他變數數據的蒐集對象一致。

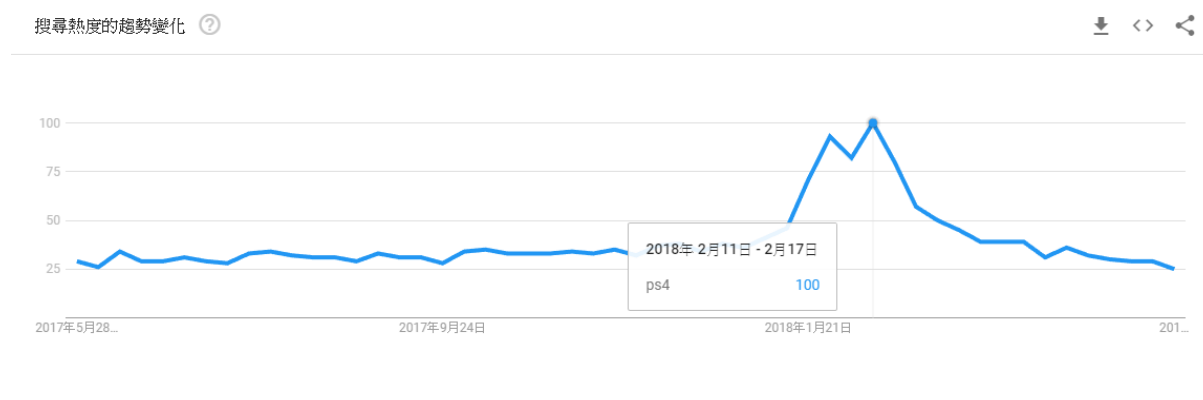


圖 10:Google Trend 示意圖

資料來源: Google Trend 取

自: <https://trends.google.com.tw/trends/explore?q=ps4&geo=TW>

另一方面，本研究也包含來自官方專頁的數據做為變數。url 為 ps4 官方專頁每一期發文中有多少比例包含 url 連結，url 連結就是一般俗稱的網址。根據 Asur & Huberman (2010) 的研究，社群網站上的發文附有 url 的比例越高，當期電影票房也越高。因此利用 url 吸引使用者去看電影的預告片或是電影官網的手法是成功的。此外，本研究也考慮 ps4 官方專頁的互動率高低的影響。互動率呈現出了用戶「參與」貼文的程度，而每個社群網站對互動率的定義不同，甚至同

一個社群網站不同時期的互動率計算公式也不相同。但只要檢視互動率，就能夠了解用戶對於貼文的反應程度。而 ps4 官方推特做為官方與玩家之間溝通聯繫的橋樑，互動率自然是重要的指標，若互動率低表示玩家可能對官方的推文不感興趣或是反映不佳，導致雖然推文的點閱率高，但是用戶卻不想在底下與官方互動，或是轉推給其他朋友看。下表 8 為變數說明。



表 8:變數說明

變數名稱	說明
銷售量 (sale)	ps4 北美區的週銷售量。數據來源:VGChartz。
發文數 (tweet)	Twitter 上蒐集到包含關鍵字的英文區每週推文數量。
PNratio	PNratio:每一期的發文經過文字情感分析後,將包含正面情感的發文數除以包含負面情感的發文數,所得的比例。
cognition	利用關鍵字搜尋每一期的推文中有多少比例屬於認知 (cognition) 期,變數值除以 tweet 即為當周比例 (cog)。推文中包含以下關鍵字則該推文屬於認知期。使用的關鍵字包含:release, publication, bestseller, hit。
affect	利用關鍵字搜尋每一期的推文中有多少比例屬於影響 (affect) 期,變數值除以 tweet 即為當周比例 (aff)。推文中包含以下關鍵字則該推文屬於影響期。使用的關鍵字包含:review, need, price, feature。
behavior	利用關鍵字搜尋每一期的推文中有多少比例屬於表現 (behavior) 期。推文中包含以下關鍵字則該推文屬於表現期,變數值除以 tweet 即為當周比例 (beh)。使用的關鍵字包含:purchase, buy, order
trend	來自 google trend 的搜尋熱度指數。在選定的時間範圍內,搜尋熱度最高的時段該數值為 100,其他時段則根據搜尋熱度的相對比例給予不同數值。
url	每週蒐集到的推文中,包含 url 連結的文章比例,公式為: $\frac{\text{包含url連結文章總數}}{\text{tweet}}$
互動率 (rate)	互動率是社群網站重要的指標之一,它呈現出了用戶參與貼文的程度。rate 為 ps4 官方專頁的每期互動率。公式為: $\frac{\text{回覆數}+\text{轉推數}+\text{like數}}{\text{追隨數}} \times 100$

資料來源:本研究整理

第五章 研究方法

第一節 模型介紹

Vapnik (1998) 所提出的支撐向量機概念，是基於 VC 維度 (Vapnik-Chervonenkis dimension) 所衍生之統計機器的學習理論 (learning machines)。

SVM (Support Vector Machine) 是一種監督式學習的演算法，試圖從資料中建構一個超平面，將資料區分成兩個類別，最後進行預測/分類 (R 筆記, 2017)。因為 SVM 考量 VC 維度中結構風險最小化原則 (structural risk minimization, SRM)，因此這個原則即處理分類問題時會尋求誤差最小化以提高分類之準確性，因而其廣泛應用於分類問題之研究。對於分類問題，SVM 就非線性之函數，將原本線性且無法分割之訓練資料 (train data) 輸入於高維度之特徵空間 (feature space)，藉由輸入資料的學習，於特徵空間中建構出可予以分類之最佳超平面 (hyper plane) 決策線，使輸入之資料達到分類效果 (董呈煌等, 2016)。其中支持向量 (Support Vector) 是指在訓練資料集 (Training data set) 中，用於分類上給予最多資訊的點，其會落在如圖 11 中所示的紅色框起來的資料點 (逍遙文工作室, 2012)。

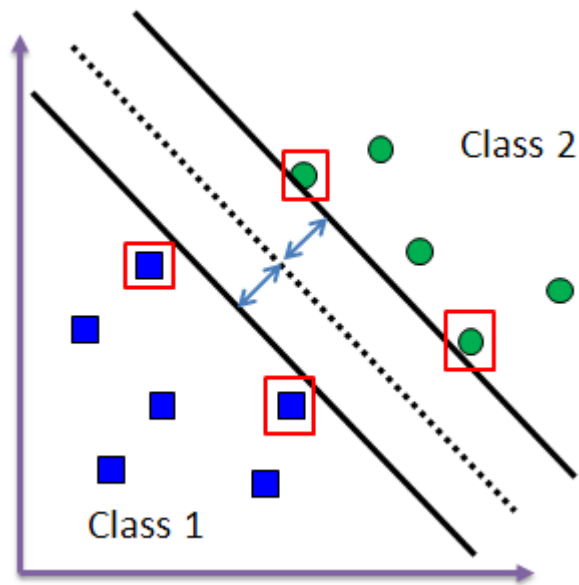


圖 11:SVM 平面示意圖

資料來源:逍遙文工作室 取自:

<https://cg2010studio.com/2012/05/20/%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%A9%9F%E5%99%A8-support-vector-machine/>

而 SVR 不同於傳統的統計模型，SVR 屬於較新類神經網路，是一種由資料驅動、無母數的模型，該模型能讓資料自己表達。相比一般的參數模型，類神經網路不需要事前確認模型的型態，且也較具有對雜訊的容忍力，在資料具有不完整或錯誤資料的情況下，也能有能力去學習這類較複雜的系統。此外，類神經網路還更具彈性，在不斷地引入新資料的情況下，它能透過不斷的重複學習而具有動態學習系統的能力。上述的這些優點就是 SVR 與傳統 OLS 統計模型的不同點 (Kaastra & Milton, 1995; Chiang et al., 1996;董呈煌等，2016)。

另外，SVM 除了可處理分類問題外，亦可擴大進行一般的迴歸預測的功能。

假設訓練資料為 $(x_1, y_1), \dots, (x_n, y_n)$ ， x_i 是投入向量， y_i 是指相對於 x_i 的產出值，SVR 是一個最適化的問題，其可表示如下式：

$$\min_{w,b,\xi,\xi^*} C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) + \frac{1}{2} w^T w \dots\dots\dots (1)$$

$$s. t. \quad w^T \varphi(x_i) + b \leq y_i + \varepsilon + \xi_i$$

$$w^T \varphi(x_i) + b \geq y(x_n) - \varepsilon - \hat{\xi}_i$$

$$\xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, \dots, N$$

其中， n 表示為樣本數， w 表示為權重向量， φ 表示為映射函數， x_i 表示為資料變數， $\varphi(x_i)$ 表示將變數 x_i 藉由函數 φ 映射至高維度的特徵空間。其中 $\xi, \hat{\xi}_i, \xi_i^*$ 為鬆弛變數 (slack variables)，目的在衡量訓練點之成本誤差。 C 為成本參數， C 越大表示容錯越小，越少的 support vectors，越逼近 hard-margin SVM 的概念，也就是能將資料完全區隔，卻也容易 overfitting。 ε 為不敏感損失函數 (insensitive loss function)，表示當實際值及預測值於容許誤差範圍內可忽略的誤差值，因此考慮實際值及預測值時， ε 則扮演極重要之權衡函數 (trade-off function)， ε 越大，代表容忍區域越大，越多資料會被忽視，造成模型的準確度降低，support vectors 的數量減少。而 ε 越小，更多的資料殘差將被考慮進去，卻也容易造成 overfitting (董呈煌等，2016；R 筆記，2017)。

1 式模型的限制條件，代表支撐向量機的過程等於解一個線性且具限制式的二次規劃問題，其所隱含的意義為 SVM 的解是唯一的、最佳化的且擺脫所得到的解為局部最小解的可能性。而 SVR 則為透過 Vapnik (1998) 的不敏感損失函數 ε (insensitive loss function) 的導入，延伸解決非線性迴歸估計的問題。根據 Vapnik (1998) 針對非線性函數，代入映射函數 φ ，將從原本非線性不可分的資

料，映射至高維度的特徵空間，以利資料在更高維度中分類，因為在高維度的特徵空間進行支撐向量迴歸之預測，將有更好的預測結果。而轉換函數通常使用核心函數(kernel function) $k(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j))$ 來替代，因可免去高維度空間複雜的計算。由於受到限制式之約束，因此導入 Lagrange 乘數 α 及 α^* 求取最佳的解：

$$\sum_{i,j} \frac{1}{2} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(X_i, X_j) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*)$$

$$\text{s.t. } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, C \geq \alpha_i^* \quad i=1, \dots, n$$

SVM 藉由訓練資料作為學習資料，將另一部分測試資料進行預測，並使用核心函數 (kernel function) 簡化資料訓練的時間。此外，資料訓練後會產生最適的核心函數，而各個資料於處理上都有屬於本身最適的核心函數，將核心函數輸入至資料中可加快分類及預測的進行，更可產生較理想的結果。在建構不同類型的 SVM 時，核心函數的選用也有不同，通常來說可分為三類核心函數：線性核心函數 (linear function)： $K(x_i, x_j) = x_i^T \times x_j$ ；多項式核心函數 (polynomial function)： $K(x_i, x_j) = [(x_i^T \times x_j) + 1]^d$ ；放射型核心函數 (radial basis function, RBF)： $K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2)$ 。(董呈煌等，2016；R 筆記，2017)

第二節 預測指標

本研究使用平均絕對百分比誤差 (mean absolute percentage error, MAPE) 以及均方根誤差 (root-mean-square error, RMSE) 作為模型預測能力的評估指標。

模型的 MAPE 與 RMSE 數值都是越小越好，越小表示該模型的預測能力較佳。其中 RMSE 的誤差權重是根據平方值，而 MAPE 是根據相對誤差。而 MAPE 各區間數值代表的模型預測能力如下表 9 所示：

表 9:MAPE 數值區間

MAPE 數值區間	預測能力強弱
MAPE 值<10	預測能力很好
10<MAPE 值<20	預測能力不錯
20<MAPE 值<50	預測能力普通
MAPE 值>50	預測能力差

資料來源:本研究整理

MAPE 與 RMSE 的公式分別如下：

$$MAPE = \frac{\sum \frac{|\text{實際值}_t - \text{預測值}_t|}{\text{實際值}_t}}{n} \times 100$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{預測值}_i - \text{實際值}_i)^2}{N}}$$

第六章 研究結果

本章節共分為三個部分，第一節包含利用 R 軟體收集到的 Twitter 數據經過整理後所繪製的視覺化圖表與敘述統計；第二節是將每一則推文內容經過斷詞處理與文本情感分析後所製成的文字雲、詞頻統計以及正負面推文統計的圖表；第三節則是以收集到的 Twitter 資料建立 SVR 模型對 ps4 銷售量進行預測，並與傳統 OLS、時間序列模型、GMM 模型以及同為機械學習演算法的隨機森林模型進行比較，並判斷哪個模型的預測能力較出色，最後作出結論。

第一節 Twitter 數據敘述統計與每週趨勢

本節為 Twitter 數據經過整理後繪製的視覺化圖表，其中包含本研究所有使用的變數，如:ps4 銷售量、推文數、PNratio、url 比例、google trend 指數、互動率，以及 AIDA 模型的三個階段的變數等。藉由這些視覺化圖表分析各個變數的每週趨勢變化情形。

表 10:敘述統計表

	Obs	Mean	Std Dev.	Min	Max
sale (單位: 34 台)	34	129271	153353.6	47961	907834
tweet (單位: 34 筆)	34	15184.76	10379.43	6426	43307
Pnratio	34	2.833	0.642	1.68	4.25
url	34	0.485	0.45	0.372	0.584
trend	34	53.15	13.89	36	100
cog	34	0.04	0.048	0.0116	0.216
aff	34	0.046	0.269	0.011	0.102
beh	34	0.037	0.16	0.016	0.1
rate	34	0.005	0.003	0.00219	0.01461

資料來源:本研究整理

首先從每週由 R 軟體所抓取到的推文總數來看，本研究由 2017 年 10 月 1 日至 2018 年 3 月 31 日為止，共抓到 30 多萬筆包含 ps4 關鍵字的推文。而每週推文數如圖 12 所示，其中在 2017/10/28、2017/11/18，與 2017/12/16 這三週的推

文數較高，都超過了 4 萬筆推文。由於每年 11 月美國的感恩節之後都會有一個黑色星期五 (Black Friday) 的特賣，ps4 的主機與遊戲在此時也會有相當優惠的折扣，因此 11 月的推文熱度也相當高。此外，12 月聖誕節也是傳統的特價促銷時段，因此在聖誕節前後的推文熱度也相當高。其他時間大致推文都落在 1 萬到 2 萬之間。

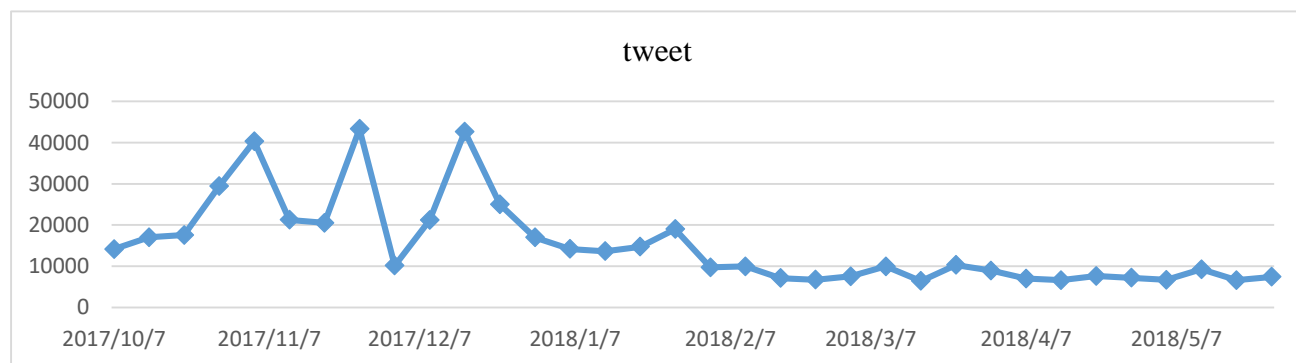


圖 12:Twitter 推文數每週趨勢

資料來源:本研究整理

再來從推文中包含 url 的比例來看，大部分的時間這個比例都在 0.45 到 0.5 之間，但黑色星期五特賣期間 (2017/11/23~2017/11/28)、2018/01/06 以及 2018/02/17 這三週的比例都較高，尤其 02/17 這週超過了 0.58；2017/12/16、2017/12/23 及 2017/12/30 這三週則較低，最低點則是 2017/12/23，僅有 0.37。若與上圖每週推文數趨勢以及下圖 PNratio 趨勢一起比較，可以發現三者都在 2017/11/25 這週都有較高的趨勢，推測是當時的促銷活動連結大量在 Twitter 上被轉發，或是在當時抽獎活動需要分享連結才能參加抽獎，因此拉高了 url 連結的比例。

根據圖 14 可以發現在 11、12 月的各週，PNratio 通常高於其他月份，或許是與這兩個月促銷特賣活動較多的關係，玩家在社群網站上的推文也較正面，PNratio 也都有 3 以上，而一月下旬是研究時段中 PN 比相對較低的時間，最低點的 2018/01/20 這週只有 1.68。

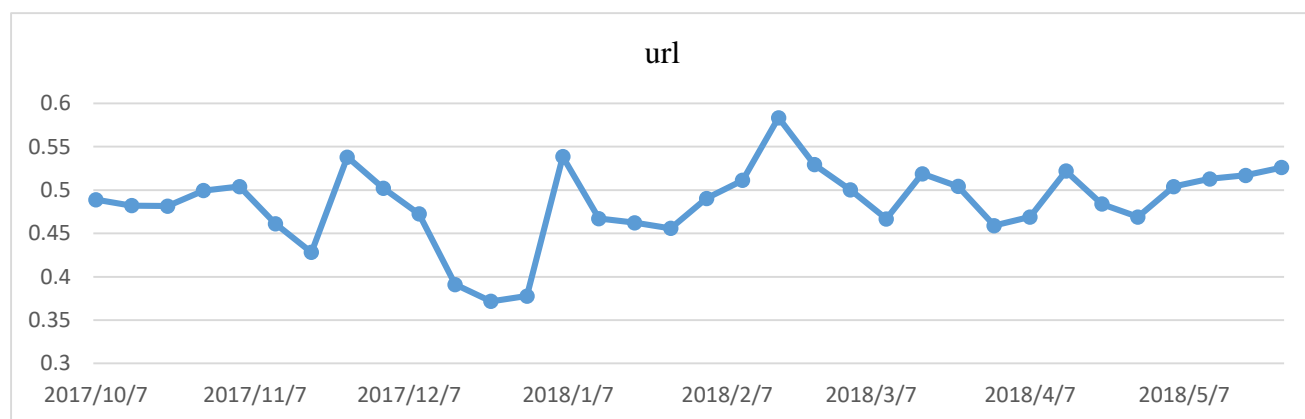


圖 13:Twitter 推文包含 url 比例的每週趨勢

資料來源:本研究整理

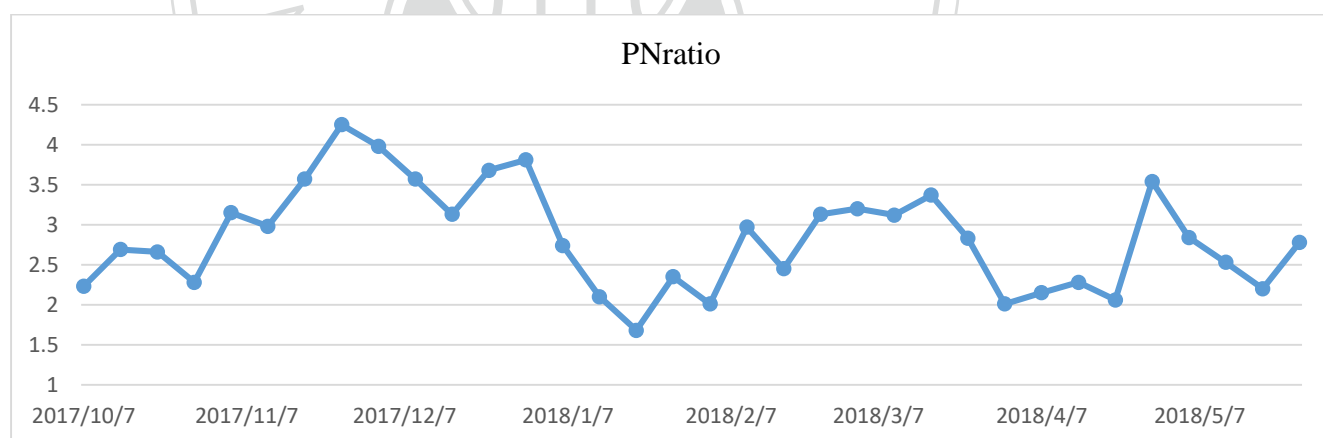


圖 14:Twitter 推文 PNratio 每週趨勢

資料來源:本研究整理

下圖 15 至 17 是代表 AIDA 模型三個階段的每週趨勢圖。由圖 15 可以看出在 2017/11/04 與 12/16 這兩週的推文中較多使用者屬於認知期，數量超過 600 則，也代表這兩週的 Twitter 用戶中有較多人開始關注 ps4 這項產品。而位於黑色星期五促銷時段的 2017/12/02 這週，顯現出使用者屬於認知期的推文是偏低的，只有 173 則，推測是在該時段大部分對 ps4 有興趣的使用者都趁特價購入，因此較少使用者處於認知期。

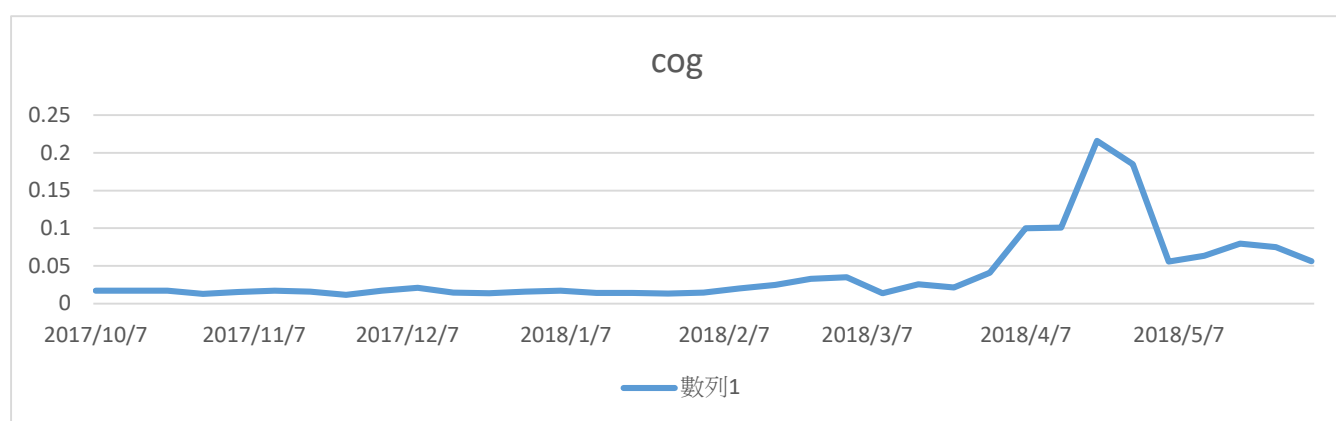


圖 15:屬於認知期的推文比例每週趨勢

資料來源:本研究整理

圖 16 是代表 AIDA 模型第二階段:屬於情感期的推文數量每週趨勢圖。與屬於認知期的推文數量的趨勢差不多，2017/11/04 與 12/16 這兩週的情感期推文數量也偏高，其中 12/16 這週更超過 1600 則，由於這時間點位於美國傳統聖誕節前夕，推測聖誕特賣可能是另一波導致 Twitter 使用者會考慮購入 ps4 的原因。而 12/02 這週是研究期間內最低的一週，僅 159 則推文表現出使用者屬於情感期。

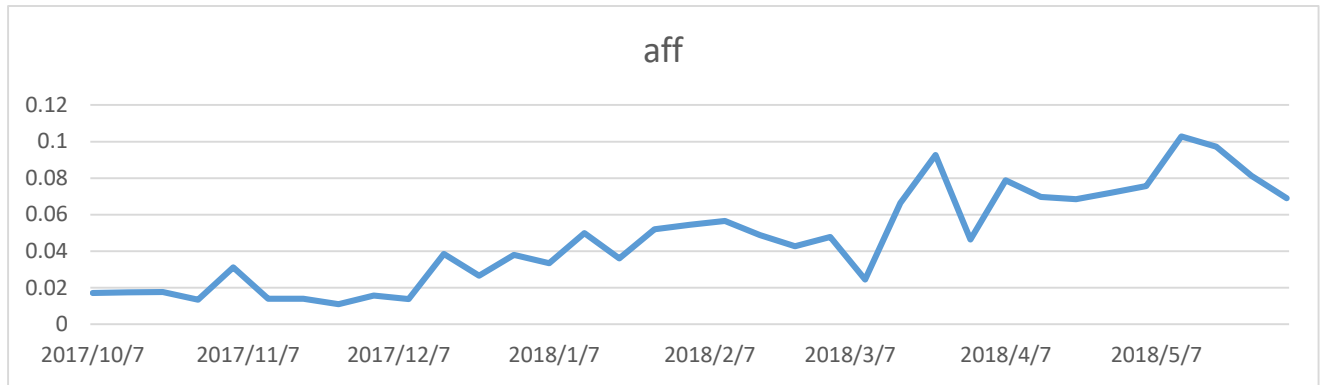


圖 16:屬於情感期的推文比例每週趨勢

資料來源:本研究整理

最後是 AIDA 模型最後階段的行為期，圖 17 顯示行為期推文數量的每週變化。由圖中可以看出在 2017/11/25 這週的行為期推文數量明顯高於其他週，接近 3000 則推文，代表有許多 Twitter 使用者在這週作出最終是否購買 ps4 的決定，而這週也剛好是黑色星期五特價開始的時間點。與圖 18 的 ps4 銷售量圖作比較也可看出在 11/25 這週的銷售量比起其他週有顯著的成長。



圖 17:屬於行為期的推文比例每週趨勢

資料來源:本研究整理

下圖 18 是美國地區 ps4 每週銷售量趨勢圖。最高點是 2017/11/25 這週，共有 907834 台 ps4 賣出，為研究期間內最高的銷售數字，這週也是黑色星期五特賣會開始的時間點。接下來的到聖誕節的幾週內，ps4 的週銷售量也都有 20 萬台上下的表現，這一個月左右的時間是 ps4 銷售量偏高的時期，到了新年過後 ps4 的銷售量以週平均約 8 萬台的速度成長。

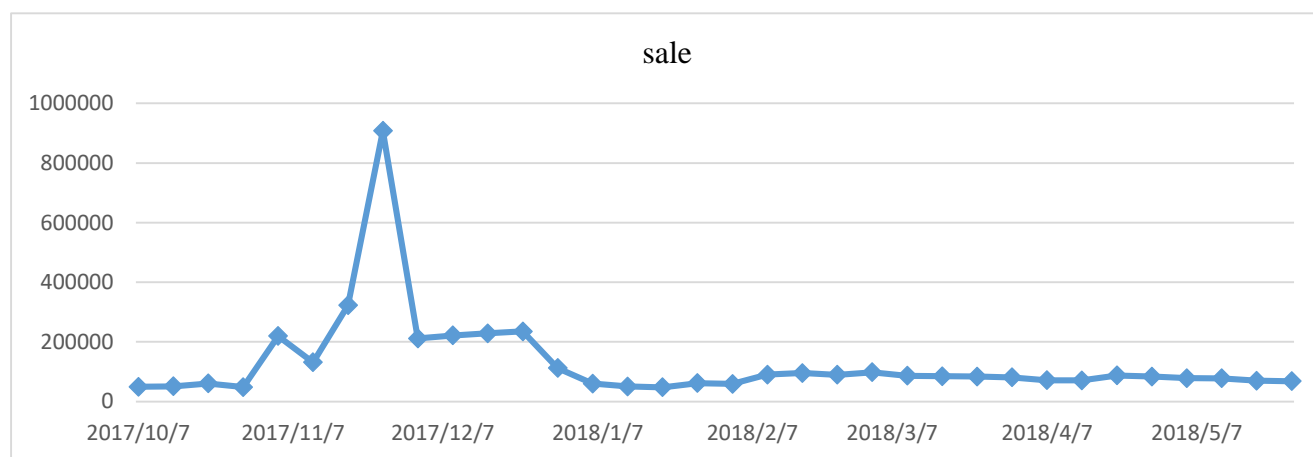


圖 18:美國地區 ps4 每週銷售量趨勢

資料來源:本研究整理

第二節 字詞分析與文字雲

本節將針對 Twitter 的推文內容進行分析，利用 R 軟體中的程式對蒐集到的推文內容進行斷詞處理，並去除無意義的字詞與標點符號之後進行統計，列出這些推文中最常使用的詞彙統計表，最後製成文字雲圖。

本研究收集的推文大多來自英語區，因此在斷詞上的處理較為方便。只要利用 R 軟體去除標點符號、超連結、http 等內容後，再依據個別需求移除一些各研

究者指定的詞彙（本研究在此刪除關鍵詞”ps4”）即可進行詞頻分析。經過斷詞處理的推文內容也將被繼續使用做為文字情感分析的素材，並製成 PNratio 折線圖（圖 14）與文字雲圖（圖 19）。

下圖 19 即為 Twitter 內容經過斷詞處理後製成的文字雲圖。可以看出在主機市場位於競爭地位的「xbox」、「one」、「xboxone」、「switch」、「nintendo」等主機與公司名稱都名列前茅，與 ps4 相關的「playstation」、「sony」也榜上有名，顯示當討論到 ps4 時，使用者們經常將這些競爭的遊戲主機擺在一起討論。而「win」、「chance」、「giveaway」、「free」這些關鍵字也大量出現因而在文字雲中較靠中間，推測這些詞可能出現在宣傳活動或是促銷的推文中，藉由抽獎獲得免費主機或遊戲軟體等手法吸引消費者目光，增加廣告文章的曝光度。為了確定這些抽獎活動與促銷活動是否能帶動 ps4 整體銷售量，本研究將以上關鍵字在推文中出現的頻率也放入模型中，P 值為 0.0146，結果顯示這些抽獎活動與促銷活動的確會使 ps4 銷售量顯著增加。再者是「pubg」、「fornite」、「controller」、「callofduty」、「cod」、「skyrim」、「fifa」等遊戲軟體或周邊相關的詞彙也出現在常用到的字詞裡，這代表當提到遊戲機時，不同公司發售的遊戲陣容也可能是重要的討論話題。最後則是與特價事件有關的字詞，例如「black」、「friday」、「christmas」、「amazon」、「bundle」顯示當 ps4 出現特價活動，該消息會透過 Twitter 的轉發快速地傳播出去，並在社群網站上被熱烈討論。

從文字雲圖與字詞統計表當中可以發現，與 ps4 相關的推文中包含了其他品牌的遊戲主機、遊戲軟體以及周邊配備，其中最常出現的是其他遊戲機的名字，位於文字雲圖正中間，其次是促銷折扣活動的情報，外圍較常出現遊戲軟體的標題。因此，玩家最在意的還是 ps4 與其他競爭對手的產品相似與相異的特色在哪裡、硬體配備如何、價格高低等，並從中決定是否購買 ps4 或是其他主機。下表

11 為本研究整理出的常用字詞詞頻統計表。



圖 19:Twitter 推文內容文字雲

資料來源:本研究整理

表 11:詞頻統計表

前 1~50 名常用字詞

xbox	one	win	giveaway	chance	get	new	amp	switch	follow
game	enter	pro	now	nintendo	play	will	edition	can	live
time	free	just	like	bundle	giving	console	gaming	retweet	playstation
winner	copy	limited	coming	code	playing	xboxone	need	fonite	good
want	best	end	sony	buy	eduaubdedubu	point	join	xboxonex	dont

前 51~100 名常用字詞

first	video	controller	christmas	pubg	back	final	black	power	check
day	hands	gameplay	available	today	away	kit	amazon	curryvip	online
know	deal	give	looking	fifa	call	anyone	running	wii	year
world	team	cant	duty	callofduty	add	update	prize	last	still
next	ever	also	news	people	entered	gamer	random	come	please

前 101~141 名常用字詞

turn	soon	tweet	night	codwii	tub	gfuel	strawberry	played	getting
custom	shocake	players	card	drawn	run	codes	release	thanks	going
supposmallstreamers	stream	brand	ive	war	cod	watch	ill	shadow	well
launch	awayfifa	walma	god	skyrim	super	toty	see	big	hunter

gamestop									
----------	--	--	--	--	--	--	--	--	--

資料來源:本研究整理

第三節 預測模型

在第三節中，本研究將以支援向量回歸 (support vector regression, SVR) 建立 ps4 銷售量的預測模型。利用 R 軟體與 package "e1071" 為工具建立 SVR 模型，並與普通最小平方法 (ordinary least square, OLS) 模型、AR(1) 以及隨機森林模型 (Random Forest) 的預測能力做比較，觀察模型間預測指標的差異並繪出 ps4 每週銷售量的預測值與實際值的比較圖。

首先建立 OLS 預測模型，結果如下圖 18 所示。由圖中可看出推特文章數量、正面情感分數與負面情感分數文章的比例 (PNratio)、含有 url 連結的推文占有所有推文比例、AIDA 模式的"表現期"以及 google trend 搜尋熱度指標的高低都與當週的 ps4 銷售量呈正相關，這樣的結果大致和 Asur & Huberman (2010) 與 Huber et al. (2011)、張哲豪 (2014) 以及 Lassen et al. (2014) 等研究結果符合。官方推特或粉絲團越頻繁地與粉絲互動，例如透過發布各式遊戲情報、商品促銷活動或是舉辦抽獎等，對產品廣告效果也會越好。而廠商若在廣告文章中加入圖片影音等外部連結，廣告效果也會比純文字更好，雖然在 Twitter 上發文有最大字數限制 (140 字)，但廠商仍應該考慮減少文字內容並將字數留給圖片海報或宣傳影片等外部連結網址以獲取更大的廣告效果，採取以圖像影音為主，文字為輔的廣告策略。PNratio 比例則代表了，當產品於社群網站中的討論評價較高且較正面時，

潛在消費者可能會因為看到社群網路上正面的心得評論而由原先的不購買或是考慮是否購買轉為決定購買，因而使銷售量提高。Google Trend 的搜尋熱度也與銷售量呈正相關，意味著許多人在購買 ps4 時都會事先透過網路搜尋相關資訊作為購買產品的考量，也印證了在網路世代消費者購買商品前的情報蒐集已經從親友介紹或是書報雜誌等管道轉為透過網路蒐集資訊與評價後再以此為依據決定是否購買該產品。而變數認知期 cog 與表現期 aff 以及互動率 rate 因影響不顯著，因此從模型中剔除。



表 12:OLS 模型估計結果

解釋變數	參數估計	P 值
tweet	7.771	0.001***
PNratio	85181.32	0.01***
url	976023.9	0.011**
trend	3520.97	0.021**
cog	454303.6	0.237
aff	-525606.6	0.515
beh	1852963	0.035**
rate	-2851036	0.657
截距	-936727	0***
F	12.44	0
Adj.R ²	0.735	

註:**與***分別代表 5%與 1%的顯著水準

資料來源:本研究整理

接著利用 R 軟體與 package "e1071" 建立 SVR 模型，本研究參考 Kohavi & John (1997) 的封套式 (wrapper) 建立模型，首先將正確率較高的變數優先選入，選入變數若對整體模型正確率有提升即保留，若正確率不變或下降則移除 (董呈煌等，2016；Kohavi & John, 1997)。經過 R 軟體與 package "e1071" 的程式碼計算後，本研究的 SVR 模型最終選入的變數有:tweet, url, Pnratio, trend, beh。而模型參數部分，在 SVR 模型中通常會調整 cost (C), epsilon (ϵ) 兩個係數以獲得更優秀的預測能力。其中 cost 決定給被誤差/分錯的資料「多少」懲罰值，C 越大則容錯率越小；C 越小則代表容錯率越大。而 Epsilon 越大，代表容忍區塊越大，越多資料會被忽視，而模型的準確度降低；Epsilon 若越小則所有資料殘差都會被考慮，但相對容易造成 overfitting (R 筆記，2017)。根據 Tay & Cao (2001) 的研究，C 係數的範圍在 1 到 100 之間時，SVR 模型可以得到較佳的預測結果 (董呈煌等，2016；Tay & Cao, 2001)。最後，根據 R 軟體套件的 Tune.model 計算之後得出本研究的 SVR 模型選擇 C=8, $\epsilon=0$ 為最佳參數組合。

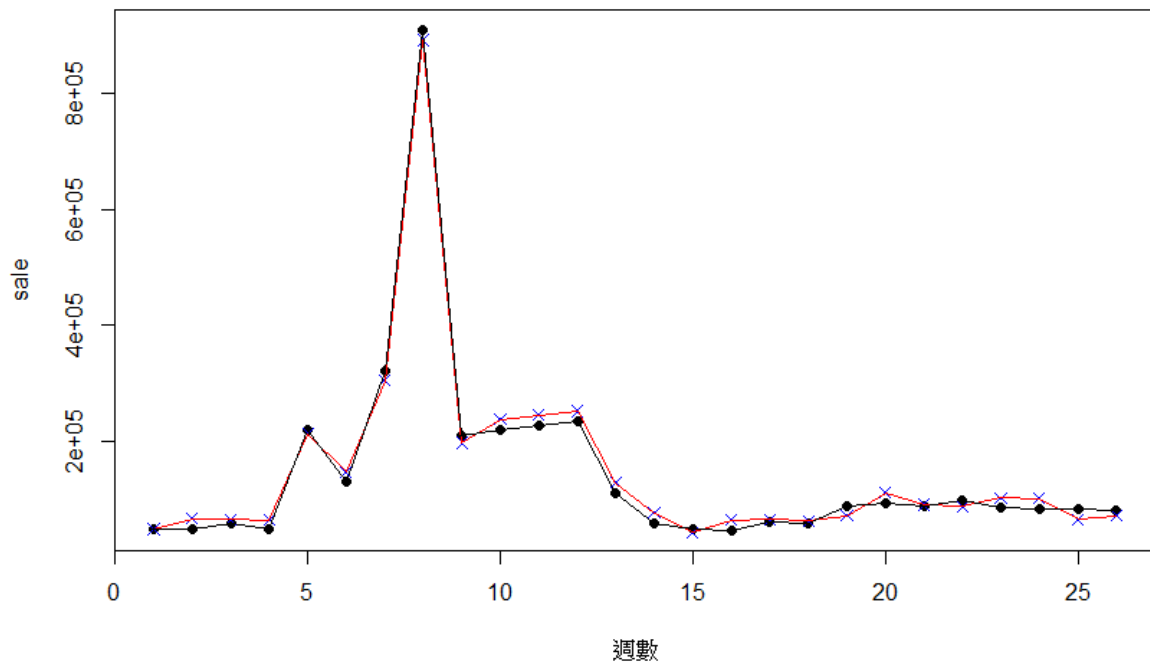


圖 20:SVR 模型的 PS4 銷售量的樣本內預測數值與實際數值折線圖

資料來源:本研究整理

接著本研究利用樣本中的前 80%資料建立 SVR 模型並進行預測，後 20%資料則用來作為與預測值對照的實際值。再將 SVR 模型與 OLS 模型、AR(1)以及隨機森林 (RandomForest) 模型進行比較，檢驗 SVR 模型的預測能力是否較優秀。分別計算出四個模型的 RMSE 與 MAPE 做為模型預測能力的指標後，從下表 13 中可以看出，不論是 RMSE 的數值或是 MAPE 的數值，OLS 與 AR(1) 的數值皆大於其他三者許多，代表兩模型預測能力差於其他三個模型，SVR 模型、GMM 模型以及 RandomForest 模型對 ps4 每週銷售量的預測數值更靠近實際數值。接著觀察 SVR 模型的 MAPE 數值約為 16.5%，對照第四章的 MAPE 預測能力區間表發現 SVR 模型屬於「預測能力不錯」的區間，略優於同為機器學習演算法的隨機森林模型以

及 OLS 與 AR(1) 模型，與 GMM 模型的預測能力則不相上下，顯示以較新的機器學習方法所建立的 SVR 模型在預測 ps4 銷售量的表現上也能達到與傳統計量模型以及傳統機械學習模型所能達成的高準確率。

表 13:SVR 與 OLS 預測效率比較表

	SVR	OLS	AR(1)	RandomForest	GMM
rmse	13225.34	38614.55	34462.26	13943.68	14114
mape	16.5%	49.23%	46.72%	18.1%	16.8%

資料來源:本研究整理



第七章 結論與建議

第一節 結論

隨著科技日新月異，民眾上網的工具推陳出新，花費在網路上的時間也增加了，俗話說「有人潮的地方就有錢潮」，在網路社群上的粉絲經營互動與消費者間的產品評價也成為廠商在宣傳商品時重要的一環。本研究利用社群網站 Twitter 取得的數據，搭配文字探勘技術分析民眾在社群網站上討論 ps4 這個主題時包含哪些方面的議題、偏好的內容，以及對 ps4 主機銷售量影響的因素，最後建立 SVR 模型進行分析預測。

本研究蒐集每周 Twitter 文章數、包含 url 連結的文章比例、google trend 搜尋熱度指標，搭配文字探勘技術的情感分析所得到的情感分數，以及關鍵字分析的文字雲、字詞統計、AIDA 模型三階段的關鍵字統計等變數，最後以支援向量迴歸 (SVR) 的方式建立一個預測 ps4 銷售量的模型。結果顯示，該模型的預測能力指標表現不錯，代表模型有能力預測 ps4 的銷售量。而 SVR 模型的預測能力表現比 OLS 出色也與之前的研究結果相同 (Vilius & Antanas, 2011; 董呈煌等, 2016)。本模型對於 ps4 銷售量影響的因素部分與先前的研究結果吻合，即說明了在這個網路蓬勃發展的時代，廠商若能在社群網站妥善經營自己的粉絲專頁、散布產品相關情報，就能夠達到廣告效果，提升產品銷售表現。

根據本研究的分析結果，推文中若是含有圖片或外部連結等 url 會比沒有包含 url 的推文得到更多關注，url 文章連結比例越高，商品的銷售量也呈正相關。因此遊戲業廠商若未來希望在社群網站上獲得更大的廣告效果，可以考慮增加 Twitter 中發文的 url 連結數量，例如附上宣傳影片、海報，或是遊戲畫面等連結。雖然在 Twitter 上發文有最大字數限制 (140 字)，但廠商仍應該考慮減少文字內

容並將字數留給圖片海報或宣傳影片等外部連結網址以獲取更大的廣告效果，採取以圖像影音為主，文字為輔的廣告策略。此外，舉辦抽獎活動也是吸引社群網站使用者瀏覽的好方法。廠商舉辦抽獎活動通常會要求參加者必須按 like 並轉推或留言，如此一來可以使該貼文快速被擴散出去，吸引更多潛在消費者來瀏覽該廠商的 Twitter 頁面或網站，並使互動率上升。

以 ps4 為例，透過字詞分析的結果可以發現抽獎活動如「giveaway」、「chance」，遊戲內容「cod」、「fortnite」，周邊配備「controller」以及與其他廠商的遊戲機「xbox」、「switch」比較的討論較容易受到關注。因此若廠商要提升官方 Twitter 的互動率以及產品銷售量可以試著從這幾點下手，例如舉辦抽獎活動、利用圖片與影像搭配文字介紹自家產品，或是比較自家產品與其他相關產品的優點，都可以吸引到更多的按讚數或是互動率。

根據迴歸分析結果可以得到幾個對 ps4 銷售量有影響的結果：正面情感的文章越多銷售量越高、搜尋熱度越高銷售量越高、url 連結比例越高銷售量越高。不只是 ps4 的銷售量受到以上因素影響，對於遊戲產業來說，例如遊戲開發廠商、周邊設備廠商，或是其他主機平台 Xbox 與 Ninterndo Switch 等遊戲相關廠商，如何利用社群網站上的廣告效果增加自家產品的銷售量都可以從以上這些方面下手。舉辦抽獎活動與圖片影音外部 url 連結的宣傳比例提高都可以有效提高文章曝光度，增加官方粉絲頁面互動率，並強化與粉絲之間的連結。例如官方 Twitter 也可以轉推其他使用者或曝光率較低的遊戲評論網站對於自家商品的使用心得與評價等文章，將更多正面的評價擴散給其他使用者，達到增加社群網站上對於該商品正面情感的討論文章數量的目的，激起更多潛在消費者的關注。

本研究藉由分析結果得到以上幾個遊戲產業能提升在社群網站中的廣告效果的因素，並建立銷售量的預測迴歸模型，不只是適用於單一項產品，其中的粉

絲專頁經營策略以及模型的各项變數都能應用於遊戲產業的其他領域，如主機製造商，遊戲開發、周邊設備製造，甚至是 VR 產業或是手機遊戲等領域都能應用。

第二節 建議、研究限制與未來研究方向

本研究因研究時間上的不足（約 8 個月）加上 Twitter API 套件對於抓取社群網站上各項資料有天數限制（7 天），因此在研究時間內所能蒐集到的樣本數偏少，在建立模型時資料庫稍嫌不足。若研究者未來在做相關研究時可以取得 Twitter premium 的帳號便可突破 Twitter 的蒐集資料天數限制，蒐集到過往的資料。除此之外也可考慮將資料蒐集的時間拉長到一年至兩年，或是將其他社群網站（如:Facebook, Instagram 等）的資料一併加入，在建立模型時或許會有更出色的結果。而其他研究成果也提到，SVR 模型在小樣本的預測表現會比大樣本更好 (Thissen et al., 2003; Zhong et al., 2009)，本研究的樣本數約 35 筆，因此可以預期 SVR 的模型表現較佳，後續研究可以大樣本的研究方向來觀察是否 SVR 模型的預測能力一樣出色。

除了家用電玩主機市場的研究，後續研究方向也可以擴大至整個遊戲產業作為研究對象，例如手機遊戲與智慧型手機等行動裝置、電腦遊戲與電腦主機市場等領域，檢視各個遊戲市場之間的社群網站廣告行銷策略是否存在差異，並可將本模型延伸或改良，成為能更普遍應用於遊戲產業的預測模型。

另一方面，本研究在認知期、影響期，以及表現期三個變數中所使用的關鍵字搜尋數目不足，沒有將所有能代表消費者處於該期的關鍵字全部納入。未來的研究可以考慮建立較完整的詞彙庫，將大部分能代表認知期、影響期、表現期的關鍵字都納入關鍵字搜尋，則能更準確地歸納出社群網路上的使用者對該產品的購買興趣究竟進展到哪一期。

而官方粉絲團的互動率這個變數，因每個社群網站計算的公式不同，對於每個變數的權重也不一樣，以 Twitter 的計算方式如下：

$$\text{互動率} = \frac{\text{回覆數} + \text{轉推數} + \text{喜歡數}}{\text{總追隨者}}$$

可以看出以 Twitter 的公式來說，一個回覆、轉推和喜歡的比重一樣，但不同類型的文章所得到的回覆、轉推，或喜歡的比例不同，即使互動率計算結果一樣，以上三個指標的分配比例可能差異很大。因此，對於在社群網站中那些文章能得到較高回覆數，或者那些文章能得到較高轉推數，而這些類型的文章又有什麼共通的特色，這些問題仍有待後續的相關研究來整理歸納。若是能找出可以吸引較多使用者回覆或轉推的文章有何共通的特色，相信對未來研究廠商在社群網站經營粉絲與廣告策略等不同目的所使用的策略這方面會更有幫助。

參考文獻

- EE Times Taiwan (2017)。2017 年穿戴式裝置出貨量估計 3.1 億台。上網日期，2018 年 6 月 11 日，檢自：<https://www.eettaiwan.com/news/article/20170829NT21-Worldwide-Wearable-Device-Sales-2017>
- Facebook. (2016).Pages Basics.上網日期，2018 年 6 月 11 日，檢自：<https://www.facebook.com/help/281592001947683>
- Global Digital Statshot Q3 2017.上網日期，2018 年 6 月 11 日，檢自：<https://www.slideshare.net/wearesocialsg/global-digital-statshot-q3-2017>。
- Google. (2016).消費者洞察報告.上網日期，2018 年 6 月 11 日，檢自：<https://apac.consumerbarometer.com/zh-tw/insights/?countryCode=TW>
- INSIDE. (2015).美國青少年最常使用社群媒體前三名：臉書、Instagram 和 Snapchat.INSIDE.上網日期，2018 年 6 月 11 日，檢自：<https://www.inside.com.tw/2015/04/28/teens-social-media-technology-overview-2015>
- R 筆記 (2017)。(14)Support Vector Machine/Regression(支持向量機 SVM)。上網日期，2018 年 6 月 11 日，檢自：<https://rpubs.com/skydome20/R-Note14-SVM-SVR>
- SONY (2017)。PlayStation VR。上網日期，2018 年 6 月 11 日，檢自：<https://asia.playstation.com/cht-hk/psvr/>
- TechNews (2016)。Canalys：VR 市場將起飛，Sony 稱王、宏達電居次。上網日期，2018 年 6 月 11 日，檢自：<http://technews.tw/2016/12/13/canalys-vr-market/>
- Wikipedia.Instagram.上網日期，2018 年 6 月 11 日，檢自：<https://zh.wikipedia.org/wiki/Instagram>
- Wikipedia.Twitter.上網日期，2018 年 6 月 11 日，檢自：<https://zh.wikipedia.org/wiki/Twitter>
- Wikipedia.虛擬實境.上網日期，2018 年 6 月 11 日，檢

自:<https://zh.wikipedia.org/wiki/%E8%99%9A%E6%8B%9F%E7%8E%B0%E5%AE%9E>

余峰璋 (2017)。臺灣數位音樂串流服務 Facebook 粉絲專頁溝通策略分析:鉅量資料分析取徑。國立臺灣師範大學大眾傳播研究所碩士論文。

吳文貴、林孟陞 (2013)。Facebook 粉絲專頁訊息廣告效果之研究：以 7-Eleven 與麥當勞粉絲專頁為例。行銷評論；10 卷 1 期 (2013 / 03 / 01)，P43 – 60。

吳佩玲 (2014)。2014 年台灣民眾之智慧化生活有感調查。資策會 FIND。

拓璞產業研究院 (2018)。2018 年 VR 市場出貨量估達 500 萬台。上網日期，2018 年 6 月 11 日，檢

自:<https://www.topology.com.tw/DataContent/release/2018%E5%B9%B4VR%E5%B8%82%E5%A0%B4%E5%87%BA%E8%B2%A8%E9%87%8F%E4%BC%B0%E9%81%94500%E8%90%AC%E5%8F%B0%EF%BC%8C%E7%8D%A8%E7%AB%8BVR%E8%A3%9D%E7%BD%AE%E7%82%BA%E7%84%A6%E9%BB%9E/455>

林永鑫、黃耀賢、林奎評、楊濠鴻 (2016)。台灣名人經營粉絲專頁成功因素之研究-以 Facebook 為例。創新與管理第十二卷第三期，25-46。

林怡靜 (2015)。台灣化妝品廠商社群媒體應用之研究。國立中央大學企業管理學系碩士論文。

林質安、劉蕙苓 (2016)。宣傳為先？互動為主？現代劇團經營 Facebook 粉絲專頁之公關溝通策略分析。藝術評論第三十期，頁 103—148(民國一〇五年)，臺北：國立臺北藝術大學。

洪培修 (2013)。入口網站橫幅廣告與社群網站廣告對品牌權益影響之比較研究。淡江大學企業管理學系碩士班學位論文。

財團法人台灣網路資訊中心 (2017)。2017 年台灣寬頻網路使用調查報告。

財團法人資訊產業工業策進會 (2017)。八成以上台灣人愛用 Facebook、Line 坐穩社群網站龍頭 1 人平均擁 4 個社群帳號 年輕人更愛 YouTube 和 IG。上網日期，2018 年 6 月 11 日，檢

自:https://www.iii.org.tw/Press/NewsDtl.aspx?nsp_sqno=1934&fm_sqno=14

高慎謙 (2017)。網路商店之網路行銷策略研究：以織襪類產品為例。公共關

係暨廣告學系(所)碩士學位論文。

張哲豪 (2014)。企業粉絲專頁之品質要素分析。品質月刊，50 (12):40-42。

陳香伶 (2014)。新聞媒體 Facebook 粉絲專頁之守門、呈現與使用——以 2013 年臺灣食用油造假事件新聞為例。國立屏東教育大學文化創意產業學系碩士論文。

董呈煌、李春長、陳俊麟、吳韻玲 (2016)。SVR 與 OLS 在住宅價格預測正確率的比較。住宅學報第二十五卷第二期，民國一〇五年十二月，學術論著，第 31 頁—51 頁。

翟本瑞 (2011)。從社區、虛擬社區到社會網絡網站：社會理論的變遷。資訊社會研究，21: 1-31。

趙滿玲 (2014)。《網路行銷特訓教材第二版》。台北：碁峯資訊股份有限公司。

數位時代 (2017)。別再錯過！VR 線下店為什麼會是台灣產業的下一個機會？上網日期，2018 年 6 月 11 日，檢自：<https://www.bnnext.com.tw/article/45930/why-vr-arcade-would-be-the-next-big-thing-for-taiwan>

鄭宇君、施旭峰 (2016)。探索 2012 台灣總統大選社交媒體之新聞來源引用。中華傳播學刊第二十九期，2016.06，109-135。

鄭宇君、陳百齡 (2014)。探索 2012 年台灣總統大選之社交媒體 浮現社群：鉅量資料分析取徑。新聞學研究第一二〇期 2014 年 7 月，頁 121-165。

鄭宇君、陳百齡 (2016)。探索線上公眾即時參與網絡化社運——以臺灣 318 運動為例。傳播研究與實踐第 6 卷第 1 期，頁 117-150，2016 年 1 月。

蕭文康 (2015)。台灣 1200 萬手機族 都愛上臉書。蘋果日報。上網日期，2018 年 6 月 11 日，檢自：<https://tw.appledaily.com/new/realtime/20151123/738735/>

謝元晟、程美華、張光昭 (2016)。運用 R 建立文字探勘平台應用於電視收視率預測。「數據分析」2016 年 6 月，11 卷 3 期，頁 109-136。

鐘慧真、梁世英譯 (2013)。Big Data 大數據的獲利模式：圖解·案例·策略·實戰。台北：經濟新潮社。原書城田真琴[2012]。《ビッグデータの衝

撃—巨大なデータが戦略を決める—》。東京：東洋経済新報社。

逍遙文工作室。支持向量機器 (Support Vector Machine)。上網日期，2018 年 6 月 11 日，檢自：

<https://cg2010studio.com/2012/05/20/%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%A9%9F%E5%99%A8-support-vector-machine/>

Social Media Lab. (2017).Facebook・Twitter・Instagram の違いと使いわけ方まとめ！特徴からユーザー動向・運用方法まで徹底解説。上網日期，2018 年 6 月 11 日，檢自：<https://gaiax-socialmedialab.jp/post-29375/>

Ayelet, G., Susan, M.Grant-Muller, Tsvi,K., Einat,M.,Silvio, N., Itay, S.. (2014).Thepotential of social media in delivering transport policy goals. Transport Policy 32 (2014),115–123.

Bamman, B., O'Connor, B., & Smith, N. . (2012).Censorship and deletion practices in Chinese social media. First Monday, 17(3).
<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3943/3169>

Belch, G. E., Belch, M. A., Kerr, G. F., & Powell, I. . (2008). Advertising and promotion: An integrated marketing communications perspective. mcgraw-hill.

Bruns, A., & Burgess, J. . (2012). Researching news discussion on Twitter: New methodologies. Journalism Studies. 13 (5), 801- 814.

David, Z., Tianjun, Fu, Xin, Li. (2009). Assessing Public Opinions Through Web 2.0: A Case Study on Wal-Mart. ICIS 2009 Proceedings.

Eszter, H.. (2015).Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Site.上網日期，2018 年 6 月 11 日，檢自：<http://journals.sagepub.com/doi/abs/10.1177/0002716215570866>

Gruzd, A., Wellman, B., & Takhteyev Y.. (2011). Imagining Twitter as an imagined community. American Behavioral Scientist, 55 (10), 1294-1318.

Jacob, G. & Ahmed, A.. (2013). Public Sentiment and Critical Framing in Social Media Content During the 2012 U.S. Presidential Campaign.Social Science Computer Review,31 (5),563-576.

Johannes, H., Andrea, L., Florian, P., Christian, R.. (2011).Stimulating User Activity

- on Company Fan Pages in Online Social Networks. ECIS 2012 Proceedings. 188.
- Kaastra, I. & S. B. Milton. (1995). Forecasting Futures Trading Volume Using Neural Networks. *The Journal of Futures Markets*. 15(8): 853-970.
- Kono, S. . (2009). From the Marketers' Perspective: The Interactive Media Situation in Japan. *Television Goes Digital*. New York: Springer. P.58.
- Kwak, H., Lee, C., Park, H., & Moon, S. , (2010). What is Twitter, a social network or a news media? Paper presented at the 19th international conference on World Wide Web.
- Lynn, Wu, Erik, B.. (2015). The Future of Prediction How Google Searches Foreshadow Housing Prices and Sales. University of Chicago Press, p. 89 – 118.
- MIT Technology Review. (2013). Language data reveals Twitter's global reach, Retrieved June 11 2018, from: <http://www.technologyreview.com/graphiti/522376/the-manytongues-of-twitter/>
- Nann, S., Krauss, J., Schoder, D.. (2014). Predictive analytics on public data-The case of stock markets. Proceedings of the 21st European Conference on Information Systems.
- Niels, B. L., Rene, M. ,Ravi,V.. (2014). Predicting iPhone Sales from iPhone Tweets. 2014 IEEE 18th International Enterprise Distributed Object Computing Conference.
- Papacharissi, Z., & de Fatima Oliveira, M. . (2012). Affective news and networked publics: The rhythms of news storytelling on #Egypt. *Journal of Communication*, 62, 266-282.
- PewResearchCenter. (2015). Social Media Update 2014, Retrieved June 11 2018, from: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf
- Sitaram, A., Bernardo, A. Huberman. (2010). Predicting the Future With Social Media.
- Techcrunch (2017) ◦ Facebook now has 2 billion monthly users... and responsibility, Retrieved June 11 2018, from: <https://techcrunch.com/2017/06/27/facebook-2->

[billion-users/](#)

Thissen, U., R. Van Brakel, A. P. De Weijer, W. J. Melssen & L. M. C. Buydens. (2003). Using Support Vector Machines for Time Series Prediction. *Chemometrics and Intelligent Laboratory Systems*. 69 (1-2): 35-49.

Thorburn, E. D. . (2014). Social media, subjectivity, and surveillance: Moving on from occupy, the rise of live streaming video. *Communication and Critical/Cultural Studies*, 11, 52-63.

VGChartz. (2015). 2015 Year on Year Sales and Market Share Update to March 21, Retrieved June 11 2018, from: <http://www.vgchartz.com/article/258786/2015-year-on-year-sales-and-market-share-update-to-march-21/>

Vilius, K. & V. Antanas. (2011). The Mass Appraisal of the Real Estate by Computational Intelligence. *Applied Soft Computing*. 11 (1): 443-448.

