

淡江大學運輸管理學系運輸科學碩士班碩士論文

指導教授：陶冶中 博士

社群媒體大數據分析應用於臺鐵安全績效與網路輿情對照
之研究

**Mapping TRA's Safety Performance with Internet Public
Opinion by Social Media Analytics**

研究生：邱柏元 撰

中華民國 110 年 6 月

致謝

光陰似箭，兩年的時間過得很快，想當初懵懵懂懂的進了淡江大學運輸管理學系運輸科學碩士班，對學校及系上老師完全不了解，直到參與新生座談時，聽了系上各位老師勉勵的話，才知道原來要走出淡江運科碩士班的大門是多麼不容易啊！雖然曾經也想過是否要放棄重考，但半途而廢不是我的作風，既然頭都已經洗一半了，那就堅持下去讀到畢業吧。如今回首過去，每天搭乘捷運淡水信義線從頭坐到尾的日子，雖然辛苦但也讓我增加更多與交通運輸工具接觸的機會，沿路上可觀察各種交通狀況，達到學以致用的目的，而不是只有死讀書。

研究所課程使我對運輸有更深入了解，但畢竟研究所要畢業，論文撰寫還是最重要的關鍵，正當我這個外系生苦惱如何找尋指導教授時，所幸得到我的恩師陶冶中老師的賞識，願意收我為陶家班門徒。陶老師不僅在運輸專業上令我折服，更是啟蒙我在運輸管理知識的一盞明燈，也教導我許多如何待人接物、做人處事的道理，很可惜只有短短的一年時間與老師在學校相處，我仍十分榮幸能受老師的指導，對老師的感激無以言表，正所謂：「一日為師，終生為父」，我永懷感恩於心中。

感謝我的另一位恩師鍾智林老師，在我尚未入學時就寫信詢問我是否願意擔任運輸工程的助教，讓我可以透過課程中吸收更多運輸有關的知識，同時給予我這個外校生很多的鼓勵與幫助，在課程安排上給了我建議，甚至在我迷失方向時引導我走向正確的道路。

感謝系上諸位老師的指導，董老師、邱老師、張老師、陳老師、劉老師、羅老師、范老師、許老師、溫老師，研究所的課程使我對運輸專業有更進一步的瞭解，在論文進度報告時的指導亦使我受益良多。也感謝玉婷助教與 Hana 助教，

剛入學時對於淡江這個環境十分陌生，系辦助教都非常貼心給予我許多幫忙，不論學務行政或系上大小事都有勞助教的處理。

感謝臺灣大學土木工程學系賴勇成教授與臺灣鐵路管理局杜微局長，特地撥冗擔任口試審查委員，惠賜指教，使學生論文更加完整。

感謝家偉學長、悅朗學長、銘倫學長、偉豪，給了我很多論文的建議，並在學科考與口試時給予我許多幫助，鼓勵我勇敢面對。

感謝東旂在我論文寫作遇到瓶頸時，總是願意耐心的聽我訴苦，分組報告時每次都竭盡所能地做到最好，讓我很放心。謝謝你這兩年的陪伴，讓我覺得備感溫暖，不管是在學校還是出去玩，有你在一起都讓我更開心，你是我最好的兄弟。

感謝研究室的同學瀧緯、皆成、恩璋、惟翔、家瑋、煜峰、凱茵、佳穎、儷媛、景勻、崇恩、禹廷、奕杉，在課業上給予我許多專業資料補充，彌補我非本系生專業知識的不足，也謝謝你們時常在研究室陪我聊天，一起交流學術議題，讓我感覺求學階段並不孤單。

最後我要感謝我的家人給予我生活上的幫助，有你們的支持和陪伴對我來說很重要，也讓我如願完成從小到大的志願，進入運輸管理學系就讀並如期畢業。此後，我將帶著各位的祝福，朝向人生的下一個目標開始邁進。

論文名稱：社群媒體大數據分析應用於臺鐵安全績效與網路輿情對照之研究

頁數：73

校系(所)組別：淡江大學 運輸管理學系 運輸科學碩士班

畢業時間及提要別：109 學年度 第 2 學期 碩士學位論文提要

研究生：邱柏元

指導教授：陶冶中 博士

論文提要內容：

近年來隨著社群媒的蓬勃發展，民眾經常使用社群媒體網路來表達自己的觀點，社群媒體儼然成為民眾發表意見、抒發情緒、彰顯自身觀點與立場的重要平臺。再加上大數據分析的流行，衍生出社群媒體挖掘技術，藉由大數據挖掘，分析民眾觀點與情感傾向，可更快速理解民眾需求。

本研究欲建立一網路輿情情感分析模式，透過蒐集社群媒體言論，運用文本分類與資料探勘技術，並將評論主題分為「車站設施」、「員工權益」、「餐飲服務」、「票證系統」、「列車運轉」五大類相關文本，分析民眾關注之臺鐵服務相關議題與輿情情感趨勢，再經由情感分析結果計算社群媒體留言之情感值，最後與臺鐵提供的安全績效(亦即事件發生資料)進行對照，以驗證本研究結果之正確性。

經由實證分析結果得知，五類評論中僅有列車運轉類別在關聯模型中最具有顯著性。平均情感分數最低的評論類別為票證系統，餐飲服務雖表現最好，但平均情感分數仍為負數，顯示臺鐵的服務仍有改善空間。

本研究借助視覺化技術，將安全績效與網路輿情兩者合併成對照圖，以時間軸檢視安全績效與網路輿情資料，探討相同時間點兩者資料的相關聯，亦即當事件或事故發生時，是否會對網路輿情的情感趨勢產生變化，以供相關單位參考之用。

關鍵字：社群媒體挖掘、情感分析、安全績效、網路輿情

*依本校個人資料管理規範，本表單各項個人資料僅作為業務處理使用，並於保存期限屆滿後，逕行銷毀。

表單編號：ATRX-Q03-001-FM030-03

Title of Thesis :

Total pages:73

***Mapping TRA's Safety Performance with Internet Public Opinion
by Social Media Analytics***

*Key word: Social media mining, Sentiment analysis, Safety performance,
Internet public opinion*

Name of Institute:

Graduate Institute of Transportation Science, Tamkang University

Graduate date: June, 2021

Degree conferred: Master Degree

Name of student: Po-Yuan Chiu

Advisor: Dr. Chi-Chung Tao

邱柏元

陶冶中 博士

Abstract:

In recent years, people have used social media networks frequently to express their opinions because of vigorous development of social media. Therefore, social media mining with big data analytics can be applied to overview public opinions and sentiment tendencies for the purpose of intelligent decision making.

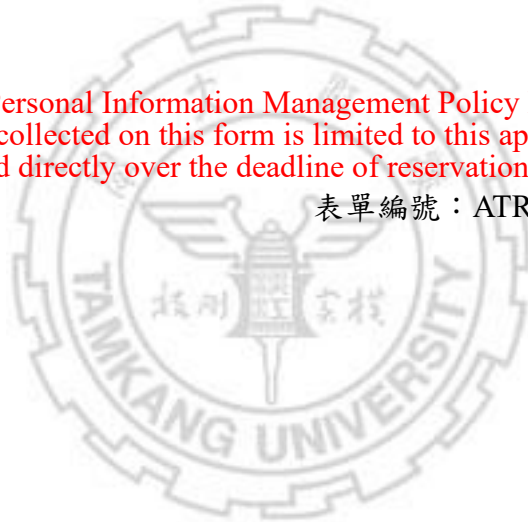
This study aims at establishing an internet public opinion analysis model with text mining technologies. Firstly, social media comments on popular websites are collected. Text classification approaches are then used to divide TRA's (Taiwan Railway Administration) service related comments into "station facilities", "employee rights", "catering services", "tickets system" and "train operation" types of topics. And then emotional values of daily comments on these five topics are calculated based on sentiment analysis. Finally, safety performance data provided by TRA (I.e. incident data) are chosen for mapping with internet public opinion results.

Empirical results showed that only "train operation" is the most significant category in the correlation with safety performance. And the lowest average sentiment score is "tickets system" category. Although "catering services" category shows positive comments, its average sentiment score is negative. It indicates that TRA's services need more active improvements.

A comparison chart can be visualized to map TRA's safety performance with internet public opinions. It is proven to be helpful for TRA's decision makers to monitor public opinion changes by using social media mining when accidents or incidents happen.

According to "TKU Personal Information Management Policy Declaration", the personal information collected on this form is limited to this application only. This form will be destroyed directly over the deadline of reservations.

表單編號：ATR-X-Q03-001-FM031-02



目 錄

第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	4
1.3 研究範圍.....	5
1.4 研究流程.....	6
第二章 文獻回顧.....	8
2.1 鐵路事故相關文獻.....	8
2.2 鐵路安全績效相關文獻.....	12
2.3 社群媒體文本挖掘相關文獻.....	14
2.4 情感分析相關文獻.....	17
2.5 文獻評析與小結.....	18
第三章 研究方法.....	20
3.1 文本挖掘.....	21
3.1.1 文本斷詞.....	22
3.1.2 文本詞向量轉換.....	24
3.1.3 機器學習.....	25
3.2 情感分析.....	29
3.2.1 情感語料庫建立.....	30
3.2.2 權重設計.....	33

3.2.3 情感辨識分類	33
3.3 安全績效與網路輿情關聯模型	34
第四章 實證分析	36
4.1 以臺鐵為本體的社群媒體資料庫建立	37
4.1.1 爬蟲系統設計	37
4.1.2 資料內容說明	38
4.2 安全績效資料	40
4.3 網路輿情資料	42
4.3.1 車站設施類別	42
4.3.2 員工權益類別	44
4.3.3 餐飲服務類別	46
4.3.4 票證系統類別	48
4.3.5 列車運轉類別	50
4.3.6 網路輿情結果分析	52
4.4 安全績效與網路輿情關聯模型	54
4.4.1 車站設施類別	54
4.4.2 員工權益類別	55
4.4.3 餐飲服務類別	56
4.4.4 票證系統類別	57
4.4.5 列車運轉類別	58
4.4.6 安全績效與網路輿情關聯模型小結	59

4.5 安全績效與網路輿情之視覺化對照圖	60
4.6 管理意涵	62
第五章 結論與建議	63
5.1 結論	63
5.2 建議	65
參考文獻.....	66



圖目錄

圖 1.1 臺鐵行車事故與事件趨勢圖	3
圖 1.2 本研究流程圖	7
圖 3.1 研究方法建構流程圖	20
圖 3.2 文本挖掘模式	21
圖 3.3 CKIPtagger 斷詞結果	23
圖 3.4 CBOW 和 Skip-gram 架構示意圖	25
圖 3.5 DNN 神經網路層架構圖	26
圖 3.6 線性整流函數 ReLu	27
圖 3.7 Softmax 激活函數	28
圖 3.8 正向情感語料庫	30
圖 3.9 負向情感語料庫	31
圖 3.10 SnowNLP 套件情感訓練過程	32
圖 3.11 情感區間尺度	33
圖 4.1 實證分析流程圖	36
圖 4.2 收集到的貼文總數占比	39
圖 4.3 收集到的留言總數占比	39
圖 4.4 臺鐵輿情文本列表範例	40
圖 4.5 行車異常事件與行車事故統計	41
圖 4.6 事故傷亡人數統計	41
圖 4.7 車站設施類別留言數	42
圖 4.8 車站設施類別情感分數	43
圖 4.9 員工權益類別留言數	44
圖 4.10 員工權益類別留言情感分數	45

圖 4.11 餐飲服務類別留言數.....	46
圖 4.12 餐飲服務類別留言情感分數	47
圖 4.13 票證系統類別留言數	48
圖 4.14 票證系統類別留言情感分數	49
圖 4.15 列車運轉類別留言數	50
圖 4.16 列車運轉類別留言情感分數	51
圖 4.17 五大評論類別情感分數分布圖	52
圖 4.18 車站設施類別情感分數與總事件數相關分析	54
圖 4.19 員工權益類別情感分數與總事件數相關分析	55
圖 4.20 餐飲服務類別情感分數與總事件數相關分析	56
圖 4.21 票證系統類別情感分數與總事件數相關分析	57
圖 4.22 列車運轉類別情感分數與總事件數相關分析	58
圖 4.23 留言情感總分與總留言數相關分析	59
圖 4.24 安全績效與網路輿情對照圖	60

表目錄

表 1.1 臺鐵事故與事件統計表	2
表 2.1 鐵路事故與事件名詞說明	9
表 3.1 推文例句訓練前後情感分數比較	32
表 4.1 爬蟲關鍵字詞庫	38
表 4.2 五大類評論結果統整	62



第一章 緒論

1.1 研究背景與動機

2018 年 10 月 21 日臺灣鐵路管理局(以下簡稱臺鐵)普悠瑪列車發生出軌事故，導致 18 人死亡，臺鐵的安全管理出現空前巨大的危機。事發後，臺鐵的安全管理機制引來外界質疑與檢討的聲浪，臺鐵亦對外宣布會深刻檢討，追究真相與責任，俾痛定思痛，大力改革，以防患未然，因此，臺鐵立即成立運輸安全處，推動安全管理系統(Safety Management System, SMS)，期能落實保障旅客安全的目標。

然而 2021 年 4 月 2 日臺鐵又發生太魯閣號出軌事故，導致 49 人死亡的慘劇！接連兩次的大型事故，民眾不滿聲浪要求臺鐵走向企業化經營，兩次出軌重大事故，更凸顯臺鐵的安全管理極需徹底整頓。

國家運輸安全調查委員會(以下簡稱運安會)歷經兩年調查，於 2020 年底公布普悠瑪事故調查的最終報告，指出四大根源錯誤，包括:組織因素、人員操作、車輛設備等長期問題，並就組織管理、維修管理、運轉管理及列車系統設計提出二十七項改善建議。然而，運安會調查的結果，卻讓民眾質疑哪一項不是人人皆知的事實？對於乘客而言，在乎的只要搭乘的列車是否能平安快速地到達目的地，此亦為運輸營運單位最該重視的一環，若缺少「安全」，其他副業做得再好，亦為徒然。

臺鐵創建至今已有一百三十多年的歷史，一直扮演著臺灣運輸交通骨幹之角色，對於國家經濟發展亦有重大之貢獻。軌道運輸在臺灣中長程大眾運輸環境占相當重要的地位，而臺鐵之發展歷史最為悠久，由於其多車種、多站等之營運特

性，列車運行環境極為複雜，若能確實檢核安全績效，制定改善或擴充策略，臺鐵安全文化才可能浴火重生。

回顧過去，臺灣發生過許多重大鐵路事故，1991 年發生造橋列車對撞事故(造成 30 死 112 傷)、2007 年發生大里列車側撞事故(造成 5 死 17 傷)，而到近年來則是 2018 年的普悠瑪新馬出軌事故(造成 18 死 215 傷)。此外，依據 2020 年政策白皮書-運輸安全顯示，臺鐵事故件數與傷亡人數並沒有逐年遞減的趨勢，表 1.1 為交通統計月報所提供臺鐵近年來行車事故與事件統計表，可以從中發現行車事件的比例仍然占大多數。

圖 1.1 為臺鐵近幾年來事故件數與傷亡人數圖，從圖中可以看出臺鐵每年行車事故與事件件數一直處於 500 件以上，且無逐年下降的趨勢，這些都顯示鐵路事故與事件在臺鐵仍需要防範發生，且 2019 年臺鐵一整年的行車事故及車輛故障次數仍高達 714 次，平均每日發生近 2 件的行車事故或故障，顯示臺鐵誤點及事故仍持續在發生，確實令人憂心。

表 1.1 臺鐵事故與事件統計表

年別(民國年)	90	91	92	93	94	95	96
事故與事件件數	857	846	718	819	877	862	835
年別(民國年)	97	98	99	100	101	102	103
事故與事件件數	825	800	853	812	639	632	590
年別(民國年)	104	105	106	107	108	109	
事故與事件件數	618	547	509	656	714	631	

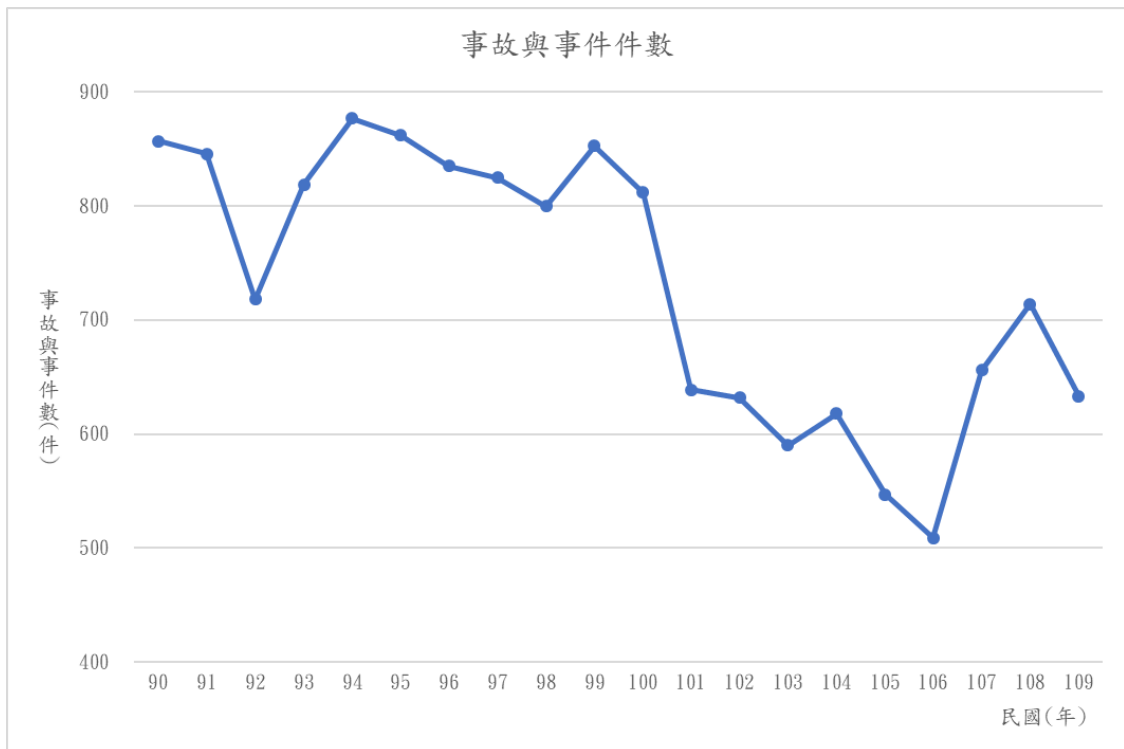


圖1.1 臺鐵行車事故與事件趨勢圖

近年來隨著社群媒體(Social Media)的蓬勃發展，民眾經常使用社群媒體網路來表達自己的觀點，每個人可以透過自己的社群媒體網頁傳佈訊息，與認識的人以更頻繁、快速且低成本的方式更新彼此的動態，社群媒體儼然成為民眾發表意見、抒發情緒、彰顯自身觀點與立場的重要平臺。因此若能針對社群媒體用戶對於臺鐵之意見評論進行挖掘，並建立蒐集、建模、分析與應用回饋之流程化作業，從中尋找利於管理階層掌握問題之關鍵所在，則可有效擬定相關政策，提升臺鐵服務品質。

有鑑於社群媒體傳播快速、內容簡短可立即撰寫、可以匿名方式發言等等的優點，因此社群媒體訊息流通的速度時常比官方機構發的公告還快速。近年來大數據分析的流行，衍生出社群媒體挖掘技術，藉由大數據挖掘，分析民眾觀點與情感傾向，可更快速理解民眾需求，以修正或改善產品與服務，進而推估輿情走向與預測未來市場趨勢。由此可知，社群媒體用戶評論的分析，無論對於產業界、學術界、政府相關單位皆具重要之意涵，若能善加分析必能從中獲得寶貴的資訊。

綜合上述，有效管理與分析網路輿情將有助於提升企業及品牌形象，透過歷史資料的研究與分析可瞭解輿情是否經妥善處理，亦可預防該事件之爆發重傷企業形象，故本研究欲建立一網路輿情情感分析模式，透過蒐集社群媒體言論，運用文本分類與資料探勘技術，分析民眾關注之臺鐵服務相關議題與輿情情感趨勢，再經由情感分析結果計算社群媒體留言之情感值，最後與臺鐵提供的安全績效(亦即事件發生資料)進行對照，以驗證本研究結果之正確性，俾利相關單位參考。

1.2 研究目的

依據上述背景與動機，本研究以臺鐵為研究對象，蒐集社群媒體中相關之評論，藉由文本資料分析，進而瞭解旅客對於臺鐵服務的滿意度與情感特徵，期望從使用者態度、顧客滿意度中更加了解使用者的需求，最後整理出安全績效與網路輿情之關聯模型，提出對臺鐵未來如何提高顧客滿意度之改善方法與建議，作為營運管理者研擬改善措施之用。研究目的如下：

1.進行社群媒體評論文本挖掘，建立臺鐵服務評論意見挖掘之分類模式。

本研究利用自行開發之爬蟲系統蒐集社群媒體文本，建立臺鐵服務評論相關的社群媒體文本資料庫，藉由資料庫當中的評論文本，可提供後續意見挖掘分類模式及情感趨勢研究使用。

2.利用資料挖掘與文本分類技術分析民眾關注話題與情感傾向趨勢，探討臺鐵網路輿情與安全績效(以事件發生為例)之對應關係，進而建立關聯模式。

本研究嘗試結合斷詞系統與自編程式而建立臺鐵網路輿情分析模式，透過斷詞系統對臺鐵網路評論文本進行斷詞，利用Python套件與自編程式完成特徵項辭庫的建置。然後針對蒐集的相關評論進行情感值計算，並根據其情

感值轉換成情感分類，再與臺鐵提供之行車事故事件統計報表進行對照，最後整理出安全績效與網路輿情關聯模型，探討兩者之關聯性。

3.針對臺鐵事件發生與網路輿情對照之結果，提出臺鐵降低事件發生之改善建議。

本研究利用臺鐵輿情分析模式獲得民眾關注話題，並將話題分類，針對不同關注話題下之網路輿情評論文本繪製時間散佈圖，並比較網路輿情發生時間與現實事件之關係，進而探討安全績效與網路輿情之關聯性，期望藉由觀察網路輿情之情感變化，了解消散之話題是否迄今仍備受關注。

1.3 研究範圍

本研究係以臺鐵營運的場站及路線為研究範圍。由於臺鐵為全臺鐵道系統歷史最悠久、營運站點最多、公里數最長、服務範圍亦最廣，因此相關評論較多，有助於後續深度學習之神經網路訓練。

首先針對與臺鐵有關之社群媒體文本進行蒐集與分析，蒐集 2018 年 1 月 1 日至 2021 年 5 月 7 日之社群媒體文本，藉以瞭解民眾對於臺鐵提供之服務看法與使用者意見情形，以社群媒體文本資料為基礎進行分析，並提出情感識別分析模型。研究範圍將評論主題分為「車站設施」、「員工權益」、「餐飲服務」、「票證系統」、「列車運轉」五大類相關文本。根據文獻回顧結果，本研究並未考慮社群媒體之傾向性，僅認為社群媒體皆多少可反映出部份民眾之真實意見。

1.4 研究流程

本研究流程為說明研究背景與動機、確立研究目的與研究範圍、回顧相關文獻、選用研究方法、進行實證分析、提出結論與建議，各階段研究步驟說明如下：

1. 確認研究目的與範圍

本研究範圍以針對臺鐵相關網路評論文本進行分析，資料蒐集時間訂為 2018 年 1 月 1 日至 2021 年 5 月 7 日止，並將蒐集到的文本資料分為五大類。

2. 文獻回顧

針對研究探索之主題進行相關文獻的蒐集與整理，文獻回顧內容包括六大部份：(1)鐵路事故；(2)鐵路安全績效；(3)社群媒體文本挖掘；(4)情感分析；最後提出文獻評析與小結。

3. 研究方法

彙整相關文獻結果，本研究應用之方法包括評論文本斷詞、評論文本轉換、情感分析、安全績效與網路輿情關聯模型。

4. 實證分析

本研究選擇 Ptt 實業坊 Railway 版上有關臺鐵的相關貼文為研究範圍，建立一套有效的自動化意見挖掘系統，用於挖掘民眾對於臺鐵服務之意見，包括資料挖掘、資料標註與預處理、模式建立與模式訓練、模式驗證，期望建立視覺化儀表板(Dashboard)，以供臺鐵管理階層決策參考之用。

5. 結論與建議

總結本研究的研究成果，提出有助於後續研究的建議。

本研究之研究流程圖如圖 1.2 所示：

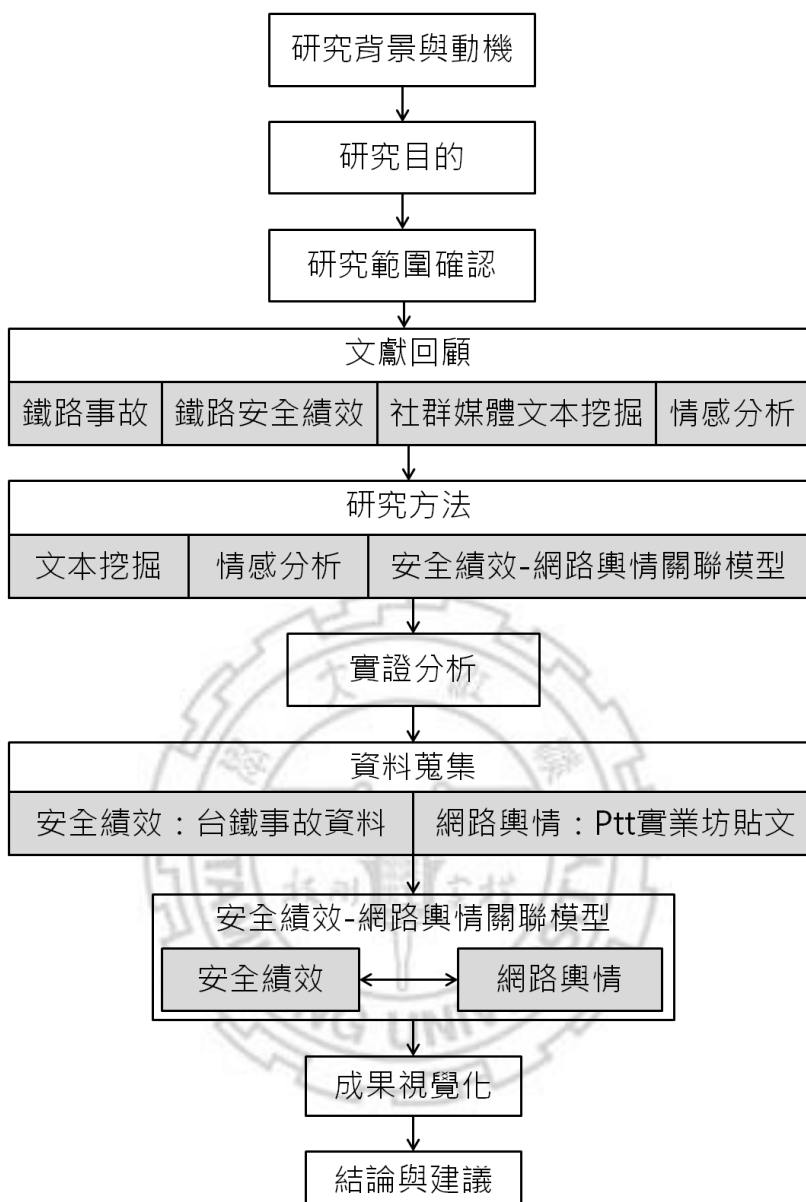


圖 1.2 本研究流程圖

第二章 文獻回顧

為深入了解安全績效與網路輿情的關聯，以及評論文本之意見挖掘所需要的技術，本研究將進行相關文獻回顧，分別是鐵路事故、鐵路安全績效、社群媒體文本挖掘、情感分析等相關文獻，從中探討前人研究成果與目前正在發展的相關文獻。

2.1 鐵路事故相關文獻

針對鐵路事故的研究相當多元，本研究所探討之鐵路事故與事件是依據鐵路行車規則第 122 條中所定義之重大行車事故、一般行車事故與鐵路行車異常事件，有關鐵路事故與事件之名詞說明詳見表 2.1。



表 2.1 鐵路事故與事件名詞說明

分類	名詞	說明
重大 行車 事故	正線衝撞事故	於正線發生列車互相、車輛互相、或列車與車輛互相間之衝撞或撞觸
	正線出軌事故	於正線發生列車或車輛傾覆或脫離軌道
	正線火災事故	列車或車輛於正線發生火災
一般 行車 事故	側線衝撞事故	於側線發生列車互相、車輛互相、或列車與車輛互相間之衝撞或撞觸
	側線出軌事故	於側線發生列車或車輛傾覆或脫離軌道
	側線火災事故	列車或車輛於側線發生火災
	平交道事故	列車或車輛於平交道與道路車輛或行人發生衝撞或碰撞
	死傷事故	列車或車輛運轉或跳、墜車致發生人員死亡或受傷之情事
	設備損害事故	列車或車輛運轉且非因天然災變造成設備或結構物新臺幣一百五十萬元以上之損害
	運轉中斷事故	列車或車輛運轉且非因天然災變造成一小時以上之運轉中斷
鐵路 行車 異常 事件	列車或車輛分離	列車或車輛非因正常作業所致之分離
	進入錯線	列車或車輛進入錯誤軌道，或於應停止運轉之工程或維修作業區間內運轉
	冒進號誌	列車或車輛停於顯示險阻號誌之號誌機內方或通過未停
	列車或車輛溜逸	列車或車輛未經駕駛員或相關人員操作控制、或錯誤操作之移動
	違反閉塞運轉	列車進入未辦理閉塞區間
	違反號誌運轉	列車或車輛未依號誌指示運轉
	號誌處理錯誤	人員錯誤操作號誌裝置或應操作而未操作
	車輛故障	車輛之動力、傳動、行走、連結、集電設備、車門、軀機、車體或其他裝置等發生故障、損壞或功能異常等影響運轉之情事
	路線障礙	土木結構物或軌道設備發生損壞、變形或功能異常致影響列車正常運轉之情事

分類	名詞	說明
	電力設備故障	變電站設備、電車線設備、電力遙控設備及其他附屬裝置等發生故障、損壞或功能異常致影響列車正常運轉之情事
	運轉保安裝置故障	列車自動控制裝置、聯鎖裝置、行車控制裝置、軌道防護裝置、轉轍裝置、列車偵測裝置、號誌顯示裝置、冒進防護裝置、災害偵測裝置及其附屬設備發生故障、損壞或功能異常致影響列車正常運轉之情事
	外物入侵	人員或外物侵入鐵路路權範圍、破壞鐵路設備、擱置障礙物或其他行為，致影響列車或車輛正常運轉之情事
	危險品洩漏	瓦斯、火藥或其他危險品從列車或車輛顯著洩漏之情事
	駕駛失能	駕駛人員於駕駛列車或車輛過程中，因身心健康因素，致無法安全駕駛或完成勤務之情事
	天然災變	強風、豪大雨、洪水、地震等其他自然異常現象，致影響列車正常運轉之情事
	列車取消	未依規定或未經核准取消時刻表訂列車班次之情事

資料來源：鐵路行車規則第 122 條

Baysari、McIntosh 與 Wilson(2008)以 19 件澳洲鐵路事故資料，透過人因分析分類法及認知錯誤的回顧和預測分析法來分析這 19 件事的人為因素，該研究發現人因分析及分類法找出的人為因素較認知錯誤的回顧和預測分析法多，而認知錯誤的回顧和預測分析法可以將人為因素更詳細的分類，雖然其無法判斷出組織的影響，但該方法可以找出重複之人為因素。

Read、Lenné 與 Moss(2012)以澳洲鐵路事故資料，透過影響因素架構圖(CFF)來分析鐵路事故，並將鐵路事故因素分為三個部分，分別是人員行為、技術失效及組織影響，並透過分群後的因素來判斷哪些為關鍵因素。

孫千山、鍾志成、李治綱、林蓁、施佑林、吳明軒、張開國、葉祖宏、賴靜慧、吳熙仁(2016)以阿里山森林鐵路出軌事故資料，透過失誤樹(FTA)、事件樹、簡易領結模型及魚骨圖方法來分析事故資料，並找出出軌之原因，也針對林務局提出軟體與硬體面的改善建議。

陳克罕(2016)以高鐵行車事件資料，經由風險矩陣研究方式，從事件發生地點、類型之分布情形及影響程度等，扣除包括颱風、地震等天然災害後，發現「道岔故障」是導致重大延誤事件發生的主要問題。透過變異數分析發現臺灣高鐵的風險事件與行車區間有關聯，其中苗栗站與臺北站發生風險事件的頻率顯著比其他車站高，再比較不同年份之間，發現不同年份之間的行車事故風險也有顯著差異。

王珊珊(2018)以美國鐵路事故資料，透過複雜網路理論與風險量化來分析事故資料，並找出事故的關鍵因素為鐵軌未固定良好與人員缺乏注意時的調車，該研究還加入動態權重，以模擬現實生活中網路的變化，進而判斷哪些節點是有關當局須注意的。

鄭永祥(2019)從普悠瑪列車脫軌事故造成重大傷亡，可發現臺鐵對於營運風險控管仍不夠嚴謹，交通部及臺鐵須以更宏觀的格局，系統性解決臺鐵的營運安全問題，可考慮導入航空領域常見的安全管理系統(SMS)，以提升臺鐵的行車安全水準，並透過積極的作為防患於未然。而政府也應協助臺鐵動態性的檢核基層人力缺口及列車、設備更新的計畫，進而讓臺鐵員工能在合理的工時下提供服務，才有可能讓員工提供安全及可靠的鐵路運輸服務。

簡信立(2020)以臺鐵行車事故紀錄進行分析，分別探討事故類型、歸咎因素對於事故發生後的影響層面，以及其交互關係。結果發現，人為疏失的出軌事故雖然佔所有事故數比例較低，但多屬嚴重以上等級事故，對於行車規則改善的可

控性較高。每件事故影響層面甚廣，最直接影響為營運業者營運成效，也可能直接或間接影響旅運乘客，因此討論人為疏失出軌事故部分與緊急應變處理之關係，找出關鍵事故原因，並就目前人為疏失出軌事故標準作業程序內容進行討論，以降低臺鐵於事故付出之成本並且加強事故排除及疏散之效率，進而提升旅客的服務品質。

2.2 鐵路安全績效相關文獻

近年來，國內鐵道業大力推廣「安全管理系統」，期望進一步提升營運安全水準。為確保安全管理系統的有效性並能持續改進，「安全保證」是相當重要的一環，可定義為檢驗 SMS 並評價其有效性的流程，包含「安全績效監測評估」、「變革管理」、「持續改進」三面向，其目的為確保風險管控措施的符合度及有效性。

李治綱、鍾志成、林杜寰、張仕龍、張恩輔、陳一昌、張開國、吳熙仁(2009)利用臺鐵 2009 到 2012 年全面性的安全實務資料進行實證分析，將事故原因分為發生於臺鐵系統內部的因素與系統外部的因素兩類，其中內部因素再分為系統(或設施)因素以及人為因素，外部因素再分為天候或環境因素以及人為因素，系統內部的人為因素，又可依照施工人員、駕駛人員、站內旅客、車上旅客等人員類型分類。最後建議臺鐵建立安全資料之蒐集、彙整、分類、分析體系，並將財物損失資料納入事故嚴重性指標，定期地、自動地與精確地進行安全自我體檢，也可以根據鐵路安全實務資料，研擬政府之長期鐵路安全政策與短期鐵路安全改善措施。

孫千山、鍾志成、李治綱、薛強、林杜寰、張仕龍、張恩輔、林蓁、黃笙玟、李永強、陳一昌、張開國、賴靜慧、吳熙仁(2011)針對鐵路運輸安全，回顧國內外鐵路風險辨識實務經驗、風險或危害項目及作法，比較分析國內外風險辨識理

論及實務，與臺鐵安全方面風險辨識實務與實作方法，包括人員、運具、場站、設施及制度等，參考國內外鐵路運輸安全辨識實務經驗與臺鐵實務作業方式，列出臺鐵安全方面所有可能的危害項目，主要區分移動、非移動、列車與天然危害四大類，可協助臺鐵進行安全風險管理，作為後續風險分析與風險評量的依據，並提出國內外鐵路運輸業蒐集分析風險辨識的作法與對策。

孫千山、鍾志成、李治綱、林杜寰、張仕龍、張恩輔、林蓁、黃笙玟、黃宏仁、張開國、賴靜慧、吳熙仁(2012)延續 2011 年「風險管理應用於鐵路運輸安全之初探-以臺鐵風險辨識為例」研究案之成果，以臺鐵為研究對象繼續進行風險管理的第二與第三個步驟－風險分析與風險評量，研究過程中回顧軌道業界常用的風險分析方法及風險評量原則，以及國內外軌道系統的實務經驗，並以「鐵公路車輛於平交道碰撞」及「旅客上下車摔倒」兩項危害為例深入探討。研究結果發現「公路駕駛因素」是平交道事故主因，其中又以「誤判短時間無車而闖越」、「未保持安全間距」、「誤判未侵入」為關鍵因素，評估本項危害屬於「必須降低」之危害，應立即採取改善手段。而「旅客上下車摔倒」危害大多肇因於人為疏失，「感知失效」、「決策失效」、「疏忽」是主要因素。

孫千山、鍾志成、李治綱、林杜寰、張仕龍、張恩輔、林蓁、黃笙玟、黃宏仁、張開國、賴靜慧、吳熙仁(2013)延續 2011 年與 2012 年研究案之成果，以臺鐵為研究對象繼續進行風險管理研究，在風險管理監督、管理改善的研究成果中，除了從國內外相關手冊與標準中彙整理論原則外，亦回顧國內外鐵道系統實務上推動風險管理之組織架構，並從中歸納七大重點工作，藉以檢視臺鐵現況風險管理推動過程中所遭逢問題，最後建議臺鐵應從組織、系統與運作等三方面進行修正，供臺鐵參考以提升整體系統安全。

林杜寰、孫千山、李治綱、陳桂豪、吳明軒、胡仲瑋、張開國、蔡祖宏、吳熙仁、洪憲忠(2019)彙整比較國外運輸業界安全管理系統做法後，提出適用我國的 12 項安全管理系統要項及作業指引，內涵 53 項檢核項目與 121 項彈性注意事項，供後續各營運機構自主盤點安全管理現況並據以研擬改善行動方案，並建議於現行法規中納入具體要求，作為鐵道營運機構必須落實 SMS 之依據，監理機構也應持續強化監理能量以確保營運機構落實安全管理系統，包括發展上位計畫、評估機制、教育訓練及細部作業規範等，透過制度化之安全管理系統(SMS)來落實風險管理，以有效避免鐵路事故發生。

徐任宏、林杜寰、孫千山、張開國、葉祖宏、吳熙仁(2021)提到安全績效指標為事故、事件的早期警訊，並可量測其輸入、輸出、結果或影響，可分為領先指標與落後指標兩類。安全績效目標為一段時間內預期的安全績效指標變化程度，須可量化並考慮其可靠度、監測數量以及是否有數據來源。營運機構必須基於國家大眾運輸安全計畫的要求項目，例如：死亡、受傷、安全事件、系統可靠度，來制定安全績效目標。

2.3 社群媒體文本挖掘相關文獻

Chan(2013)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(2013) 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」。

游綉雯(2015)使用 TripAdvisor 線上評論資料作情緒分析，計算出情緒分數，透過迴歸分析了解其相關性，並將評論分類了解消費者滿意不滿意之項目。研究

證實情緒分數對於滿意度具有效度，情緒分數可代表消費者對企業的產品或服務之滿意度。

朱斌妤(2015)蒐集網路上有關自由經濟示範區的討論文章，並篩選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，觀察網路輿情在事件發生前後的變動情形。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，因此建議進行輿情分析時，若要了解短時間變化宜採大數據方法；考量民眾深度想法則可採取問卷調查之策略。

陶冶中與簡睿志(2016)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，提供給政府推動多元化計程車服務之參考。結果顯示，民眾對於UBER在臺營運之網路平臺服務的評價，整體情感趨勢係以正面偏中立為主，高於負面情感傾向。

陳翰(2018)針對民眾日常交通進行社群媒體文本挖掘。首先建置社群媒體文本資料庫，並將日常交通之社群媒體文本進行情感分析，以瞭解民眾對於各公共運具評論之情緒，接著使用深度學習CNN演算法建構多元情緒決策情感分析模型，再經由實證分析，運用文本挖掘及深度學習分類技術，分析民眾日常通勤通學搭乘公共運具之滿意度，並以五個評級界定(非常滿意、很滿意、普通滿意、不滿意、非常不滿意)。

張悅朗(2020)利用深度學習演算法，建立以軌道運輸場站為本體的社群媒體資料庫，分析評論文本與圖像，並將評論分析結果結合地理位置、時間因素，以視覺化圖形呈現，供管理者檢視。透過實測分類結果，建構軌道運輸場站服務評論意見挖掘之分類模式，並以軌道運輸場站為核心，探討意見挖掘結果與實際狀況之關係。

徐倩文、韋健華、吳忠宏、董二為(2020)採用 Kaplan-Meier 法和 Cox 迴歸模型，探討在網路上關於旅遊負面事件的網路輿情生存時間。研究結果顯示，事件發生後的應對時間與事件生存時間成反比，官方主體回應會對事件熱度生存時間產生顯著影響，採取解釋說明的應對措施，能夠有利於縮短輿情事件的生存時間。然而，負面事件的應對一方面延長了討論熱度生存時間，另一方是負面事件補救效應的傳遞，從負面事件網路熱度的生存時間與應對時間成反比的結論可以看出，當事件做出應對後網路熱度迅速衰減，因此負面事件的快速應對更有利於降低其負面影響。

Arjona(2021)從馬德里地鐵的 Twitter 帳號收集兩個月的數據，使用文本挖掘和機器學習演算法識別評論的主題和情緒，以探索評論者的空間與時間分布，最後再與官方做結合，利用地理加權回歸模型來探索投訴評論者的空間分佈。結果顯示，Twitter 使用者往往是居住在地鐵周邊地區的中等收入員工，在前往工作場所時透過推特發文，並發現地鐵的主要問題是集中在轉運站及市中心地區，像是誤點和故障，主要是在平日清晨，且受目的地區域景點密度的影響。

Rachman(2021)透過 Twitter 的數據，試圖將民眾觀點歸類為正面和負面來分析民眾對公共交通的看法，並使用 Latent Dirichlet 分配方法 (LDA) 進行評論主題建模，以找出每個公共交通工具經常被公開討論的主題。研究發現，KRL 是最受負面情緒影響的公共交通類型，此外，對於捷運和輕軌的評論傾向於提及具有正面情緒的用詞。在民眾輿論的分群中，LDA 模型產生了關於公共交通的討論主題，該話題針對每種公共交通類型而有所不同，例如：關於 Transjakarta 線路擁塞的討論、KRL 列車的延誤、地鐵站的垃圾數量、認為錯誤開發的輕軌等等。

2.4 情感分析相關文獻

情感分析一詞最早源自於 Das 與 Chen(2001)對股票市場留言版之文本訊息，將訊息中蘊含情感定義為的正面(positive)與負面(negative)意見。Dave(2003)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。

陳亭愷(2015)應用國道電子收費政策議題，進行網路意見文本挖掘，利用電腦辨識內容後進行詞率計算找出關鍵詞，再透過爬蟲獲取文本、比對語斷詞、給予權重後，接著利用文本分類演算法將文本加以分類賦予情感值，賦予文本評論一個數值化的情感值，以正負數值表示正情感、負情感，瞭解民眾網路討論話題之情感傾向趨勢。

蔡易辰(2016)依照三元決策理論以及情感傾向分析方法，建構出三元決策情感分析模型，並將其使用在國道計程電子收費之網路輿情文本與遠通電收 ETC 客服語音資料。所建構之三元決策情感分析模型藉由實證分析證明民眾對於遠通電收公司之企業形象不僅具有正面、負面情感，亦具有中立情感的存在。

陶冶中與簡睿志(2016)提到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：文本的主客觀分類、文本的極性分類、判定文本情感傾向的強弱程度。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；文本的極性分類(polarity classification)，又稱為正負向情感分類，即判斷文本內容是正面的肯定讚賞或是負面的否定批判；判定文本情感傾向的強弱程度，如：非常不滿意、不滿意、中立、滿意、非常滿意等，又稱為等級推理(rating inference)。

洪敬政(2017)提到情感分析是辨識人類情緒及表達的系統化分析方法，網路上大量的情感分析資訊目前已能透過解構並提取出有意義價值的參考資訊。研究證實，透過社會網絡分析呈現情感分析研究之可視化關係圖，並發現「分類」為情感分析研究領域的重要主軸。

趙玉娟與陶振超(2018)提到情感分析是一種以電腦運算方式衡量文本的意見傾向與情緒反應的方法，情感分析的發展的重點在於能準確地判斷網路留言的情緒反應。其判別方法可分為兩大類：正負面情緒與 N 點量表。

2.5 文獻評析與小結

從鐵路事故相關文獻回顧中，可了解到以往研究鐵路事故因素大都採用風險矩陣、複雜網路理論與風險量化來分析事故資料。根據交通部運輸研究所及多篇對臺鐵事故的研究，皆顯示臺鐵對於營運風險控管仍不夠嚴謹，欲解決臺鐵的營運安全問題，可考慮導入航空領域常見的安全管理系統(SMS)。找出關鍵事故原因，降低臺鐵於事故付出之成本並且加強事故排除及疏散之效率，進而提升旅客的服務品質。

從鐵路安全績效相關文獻回顧中，可了解到安全績效指標為事故、事件的早期警訊，並可量測其輸入、輸出、結果或影響，安全績效目標為一段時間內預期的安全績效指標變化程度，須可量化並考慮其可靠度、監測數量以及是否有數據來源。本研究參考相關文獻，將事故與事件發生次數定為本研究之安全績效指標。

從社群媒體圖文挖掘相關文獻回顧中，可了解到社群媒體數據能提供有關服務使用和體驗的資訊，把性能指標與用戶對服務的期望進行整合，透過情感分析可以了解使用者對於評論之情緒。且與情事件發生後的應對時間與事件生存時間成反比，官方主體回應會對事件熱度生存時間產生顯著影響，負面事件的快速應對能夠降低其負面影響。

從情感分析相關文獻回顧中，可了解情感分析是辨識人類情緒及表達的系統化分析方法，網路上大量的情感分析資訊目前已能透過解構並提取出有意義價值的參考資訊，其判別方法可分為兩大類，分別是正負面情緒與 N 點量表。藉由情感分析能了解民眾網路討論話題之情感傾向趨勢，意見挖掘與情感分析的結合，應用於交通運輸領域，具有研究價值。

有鑑於此，本研究擬以陳亭愷(2015)、蔡易辰(2016)、陳翰(2018)與張悅朗(2020)等碩士論文所累積之理論與實際工具為基礎，導入系統績效(System Performance)與使用者體驗(User Experience)對照之關聯概念，針對臺鐵安全績效資料之時間序列變化對照民眾關注話題下之網路輿情評論文本而繪製時間散佈圖，並對照網路輿情發生時間與現實事件之關係，探討安全績效與網路輿情之關聯性，期望可藉由觀察網路輿情之情感變化，探討所關注安全績效話題之消散情形。



第三章 研究方法

本研究目的是建立臺鐵網路輿情意見挖掘之分類模式，利用社群媒體挖掘臺鐵旅客體驗相關的輿情文本，再與臺鐵提供的行車事故報表資料搭配，探討民眾對於臺鐵服務的體驗情感特徵與評論意見內涵。

從文獻回顧得知，本研究將蒐集評論文本、評論文本預處理、評論文本轉換、評論分類輸出，最後進行情感值計算。過程中需將安全績效與網路輿情資料各自處理，再統整成安全績效與網路輿情關聯模型。本研究之研究方法流程如圖 3.1 所示。

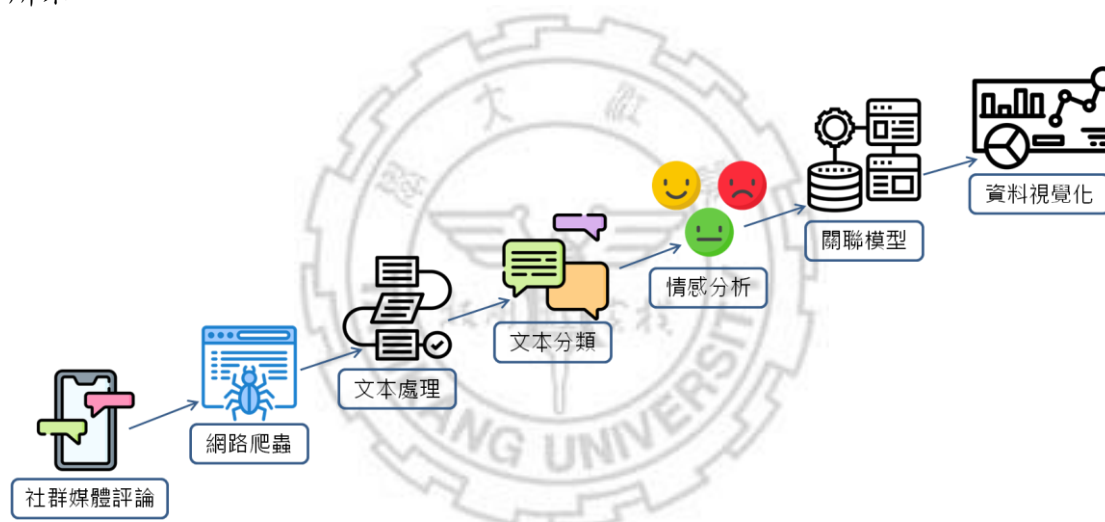


圖 3.1 研究方法建構流程圖

本研究將針對應用之研究方法，包括評論文本挖掘、情感分析、安全績效與網路輿情關聯模型等，分別說明如下：

3.1 文本挖掘

文本挖掘係從非結構化文本資訊中擷取使用者感興趣或有用模式的過程。應用知識挖掘技術，對網路上的數據文本進行蒐集，並利用電腦系統辨識內容後加以歸類、分析，以獲得進一步的研究結果。一般而言，文本挖掘技術可概略分為文本蒐集、文本辨識、文本分類等三種，如圖 3.2 所示。

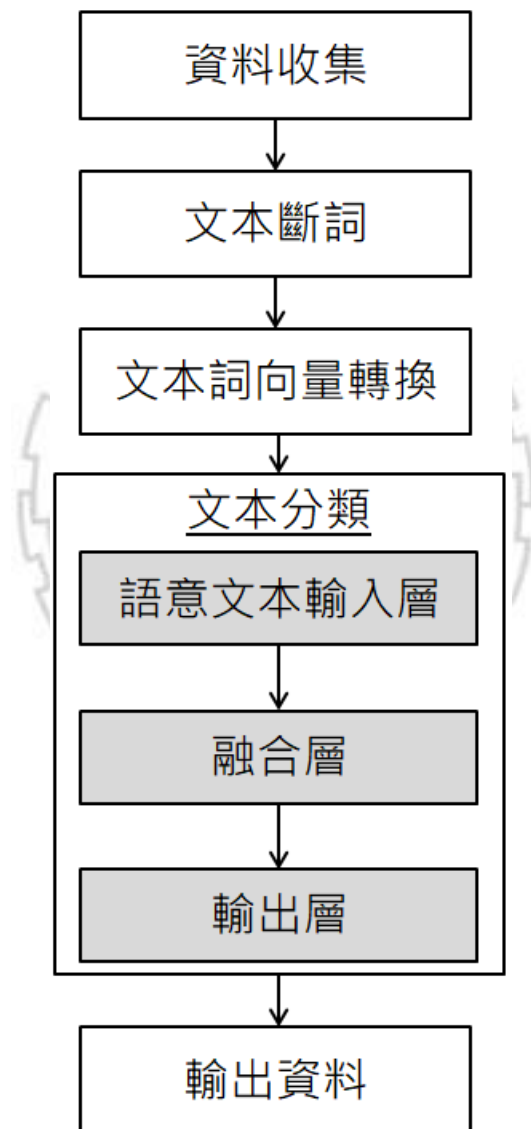


圖 3.2 文本挖掘模式

3.1.1 文本斷詞

中文的詞是最小有意義且可以自由使用的語言單位。任何語言處理的系統都必須先能分辨文本中的詞才能進行進一步的處理，例如：機器翻譯、語言分析、語言了解、資訊抽取等等，因此中文自動分詞的工作成為語言處理不可或缺的技术。自動分詞多利用資料庫詞典中收錄的詞和文本進行比對，找出可能包含的詞，由於存在歧義的切分結果，因此多數的中文分詞程式多討論如何解決分詞歧義的問題，而較少討論如何處理詞典中未收錄的詞出現的問題。

然而中文的自然語言處理較為特殊，其中之一是連續的中文文本句子，舉例「太魯閣/出軌/釀/50/死/202/傷」一句，可看出中文句子並不像英文句子字詞之間存在空白分隔。因此，進行中文自然語言處理之前，必須先分辨出文中的句子構造而進行斷詞，才能準確辨識文本語意。針對中文自然語言處理的斷詞處理，目前較流行的工具有開源函式 CkipTagger、Jieba 等。

CKIPtagger 中文斷詞套件是我國中央研究院的中文詞知識庫小組於 2019 年採用 GPL 3.0 協議開源釋出在 GitHub 平臺上的中文斷詞工具。實際上該工具已運作多年，但過去因須連接網路才可使用斷詞功能，因此造成不便而未能廣泛流行，現在開源化的 CKIPtagger 推出後，開發者可線下使用該系統，依據自己需求來處理文本、語義分析，包括更改權重、加入自訂斷詞字典等。CKIPtagger 斷詞主要是結合規則斷詞與統計斷詞兩種策略，規則斷詞是透過詞典，在對句子進行斷詞的時候，將句子的每個字與詞典中的詞去進行匹配，若找到則斷詞，否則無法斷詞；統計斷詞是看相連的字在不同的文本中出現的次數越多，就推斷這幾個相連的字很可能就是一個詞，因此可以利用字與字相鄰出現的頻率來做統計，當高於某一個臨界值時，便可認定此字組為一個詞語。

CKIP 斷詞系統是根據中研院自行設計的深度學習演算法，特色在於可將文本依詞性進行斷詞，並呈現斷詞、詞性以及實體辨識之結果，除了斷詞與詞性標記外，專有名詞辨識，或稱實體辨識 (Named Entity Recognition, NER) 是非常實用的功能，其目標為在文字資料當中，能夠辨識出感興趣的專有名詞(包含原本資料庫不存在的新專有名詞)，並自動標記正確的分類，如人名、地名、組織名等等，是人工智慧當中理解語言的重要步驟，如圖 3.3 所示。

'太魯閣出軌釀50死、202傷，交通部長林佳龍已向總統府、行政院口頭請辭，林佳龍今(4/4)白天受訪雖未親口對媒體證實，不過林佳龍今晚在臉書上表示，得知消息後他祈求上天讓乘客平安獲救，但每傳來一個壞消息，他內心更自責內疚，林佳龍晚間也在臉書上正式宣布辭去交通部長一職。'

太魯閣(Nc) 出軌(VH) 釀(VC) 50(Neu) 死(VH) 、(PAUSECATEGORY) 202(Neu) 傷(VH) ，(COMMACATEGORY) 交通部長(Na) 林佳龍(Nb) 已(D) 向(P) 總統府(Nc) 、(PAUSECATEGORY) 行政院(Nc) 口頭(D) 請辭(VA) ，(COMMACATEGORY) 林佳龍(Nb) 今(Nd) ((PARENTHESISCATEGORY) 4/4(Nd))(PARENTHESISCATEGORY) 白天(Nd) 受(P) 訪(VC) 雖(Cbb) 未(D) 親口(D) 對(P) 媒體(Na) 證實(VE) ，(COMMACATEGORY) 不過(Cbb) 林佳龍(Nb) 今晚(Nd) 在(P) 臉書(Na) 上(Ng) 表示(VE) ，(COMMACATEGORY) 得知(VK) 消息(Na) 後(Ng) 他(Nh) 祈求(VE) 上天(Na) 讓(VL) 乘客(Na) 平安(VH) 獲救(VA) ，(COMMACATEGORY) 但(Cbb) 每(D) 傳來(VC) 一(Neu) 個(Nf) 壞(VHC) 消息(Na) ，(COMMACATEGORY) 他(Nh) 內心(Na) 更(Df) 自責(VA) 內疚(VH) ，(COMMACATEGORY) 林佳龍(Nb) 晚間(Nd) 也(D) 在(P) 臉書(Na) 上(Ncd) 正式(VH) 宣布(VE) 辭去(VC) 交通部長(Na) 一(Neu) 職(Na) 。(PERIODCATEGORY)

(0, 3, 'FAC', '太魯閣')

(6, 8, 'CARDINAL', '50')

(10, 13, 'CARDINAL', '202')

(15, 19, 'ORG', '交通部長')

(19, 22, 'PERSON', '林佳龍')

(24, 27, 'ORG', '總統府')

(28, 31, 'ORG', '行政院')

(36, 39, 'PERSON', '林佳龍')

(45, 47, 'TIME', '白天')

(61, 64, 'PERSON', '林佳龍')

(64, 66, 'TIME', '今晚')

(95, 96, 'CARDINAL', '一')

(110, 113, 'PERSON', '林佳龍')

(113, 115, 'TIME', '晚間')

(126, 130, 'ORG', '交通部長')

圖 3.3 CKIPtagger 斷詞結果

本研究將使用 CKIPtagger 作為社群媒體評論文本的斷詞工具，按評論句子結構逐項分開，即主詞、動詞、賓語等分開，並納入由本研究自訂的鐵道版相關詞彙的字典和權重，以獲取更理想的斷詞結果。

3.1.2 文本詞向量轉換

文字型態的資料，是沒有辦法如同數值資料一樣進行 Forward Propagation 以及 Backward Propagation 的，簡單來講，是沒辦法進行運算並且訓練權重網路的。為了要解決這個問題，需要將「文字」轉換成「數值」，最簡單的方法，就像是將每個相異字(Character)轉換成一個特定的數字，這樣一來，便可以將機器學習相關的技術應用在文字領域。但只是單純地轉換 Character，有時是得不到較好的結果的，對中文而言，使用「詞」作為句子裡的基本元件單位會更恰當。另外，只是轉成一個「數字」也很難表現出中文詞彙的多樣性，轉換成「向量」通常效果會更好一點。

文本需要提取特徵值才能讀取至電腦運算，詞向量是轉換評論文本至特徵值的常見方法。詞向量是其中一種將資料中的單詞透過學習訓練轉化為向量表示的一種技術。詞向量通常是密集(dense)、實值(real value)，詞向量的每個維度是用來表示每個字詞的各個語義(semantics)上的潛在特徵，不少學者在過去提出不同的演算法來提取詞向量，例如 Softmax、LBL 模型、Word2vec、Glove 等。

Word2Vec 模型就是這樣的一個工具，其特色是透過 Hierarchical Softmax 將訓練資料中不同字詞建置在霍夫曼樹(Huffman tree)上進行最佳化，另外以負面採樣(negative sampling) 取代最大似然估計(MLE)，大幅加速其運算速度。Word2vec 有兩種模型架構—CBOW (Continuous Bag-of-Words Model) 和 Skip-gram(Continuous Skip-gram Model)，如圖 3.4 所示。CBOW 是利用上下文來預測中心詞，Skip-gram 則是利用中心詞來預測上下文。Skip-gram 嘗試根據同一句子中的另一個單詞來最大化單詞的分類，即每個單詞會被輸入至連續投影層的線性對數分類器(log-linear classifier)，以預測每個單詞前後某個範圍內的單詞出現機率。

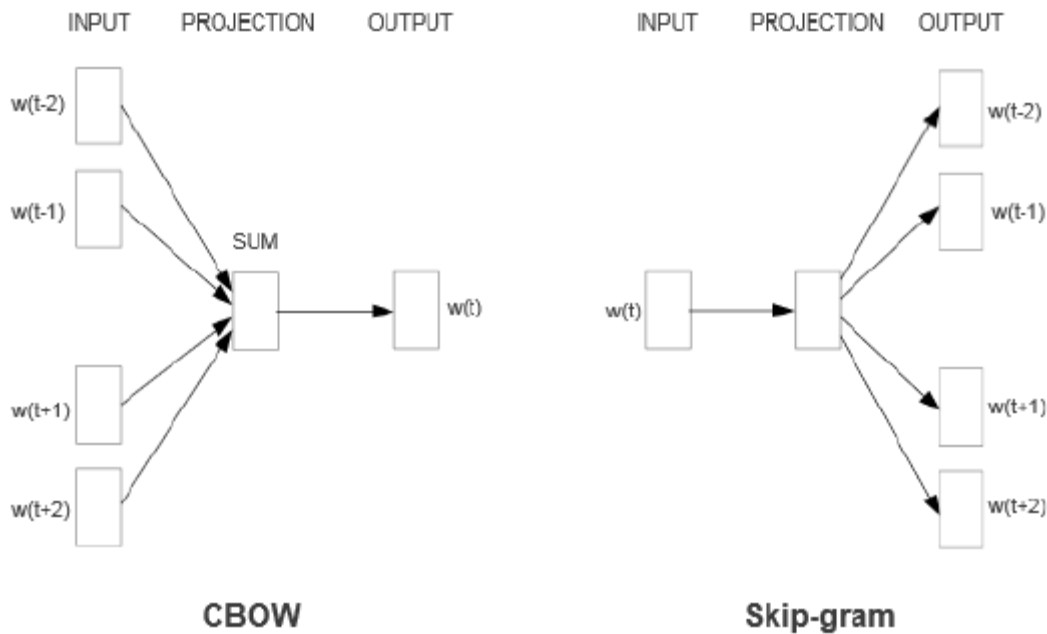


圖 3.4 CBOW 和 Skip-gram 架構示意圖

3.1.3 機器學習

深度學習是由神經網路(Neural Network)演變而來，通常指具有多個網路層所組成的神經網路，透過訓練學習資料特徵，以實行自動化判斷。深度學習可分成兩個階段，分別是「學習與訓練」和「推論與預測」，前者經由機器學習技術，利用大量樣本數據，並對演算法進行訓練；後者則是執行演算法，在終端應用之處，解讀並辨識真實世界的數據，常見的深度學習演算法包括深度神經網路 DNN (Deep Neural Network)、卷積神經網路 CNN (Convolution Neural Networks)、遞迴神經網路 RNN (Recurrent neural network)，皆為深度學習中最高知名且最熱門的技術。以下將說明本研究使用的神經網路常用組件、文本處理之 DNN 演算法架構。

DNN 深度神經網路，是機器學習中的一種深度學習方法，透過仿神經系統的數學模型，進行多次不同階層與架構的運算和訓練，找出最有效、最佳化的深度學習模型。和過去機器學習需要由人來提供規則的方式相較，DNN 只需要在

設計好的神經網路中，藉由訓練資料與參數的微調設定，機器就能自行學習、找出特徵，算出最好結果。

DNN 內部的神經網路層可分為三層，輸入層、隱藏層和輸出層，如下圖 3.5 所示，最左邊的是輸入層，中間的層數都是隱藏層，而最右邊是輸出層。給定節點獲取其輸入的加權和，並將其傳遞給非線性激活函數，這是節點的輸出，然後成為下一層中另一個節點的輸入，訊號從左向右流動，最終輸出通過對所有節點執行此過程來計算。

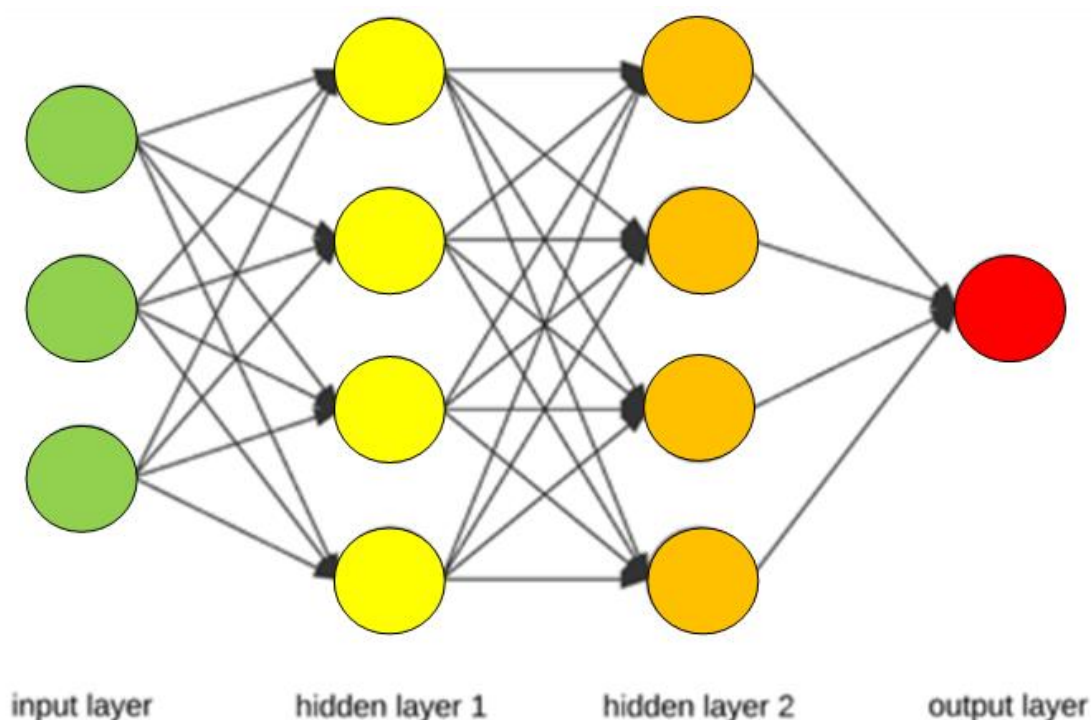


圖 3.5 DNN 神經網路層架構圖

輸入層(Input layer)是神經網路的起始，處理資料集的輸入，以矩陣形式傳遞至其他網路層。本研究模式為語意文本輸入。

隱藏層(Hidden layer)可以有許多層，以增強模型的表達能力，層與層之間是全連接的，也就是說，第 n 層的任意一個神經元一定與第 $n+1$ 層的任意一個神經元相連。雖然 DNN 看起來很複雜，但從小的局部模型來說，還是和感知器一樣為一個線性關係。

激活函數(Activation function)存在於各隱藏層、輸出層，其功能是使神經網路能夠以非線性的方法來解決非線性的問題，少了激活函數就會變成線性處理方式。在類神經網路中如果不使用激勵函數，那麼在類神經網路中皆是以上層輸入的線性組合作為這一層的輸出（也就是矩陣相乘），輸出和輸入依然脫離不了線性關係，做深度類神經網路便失去意義。常見激活函數包括 Sigmoid、tanh、線性整流函數(Rectified Linear Unit, ReLU)。本研究使用 ReLU 作為隱藏層激活函數，ReLU 已被廣泛使用，其優點在於收斂速度快、不容易發生梯度消失問題及計算資源使用較低。ReLU 圖形如圖 3.6 所示。

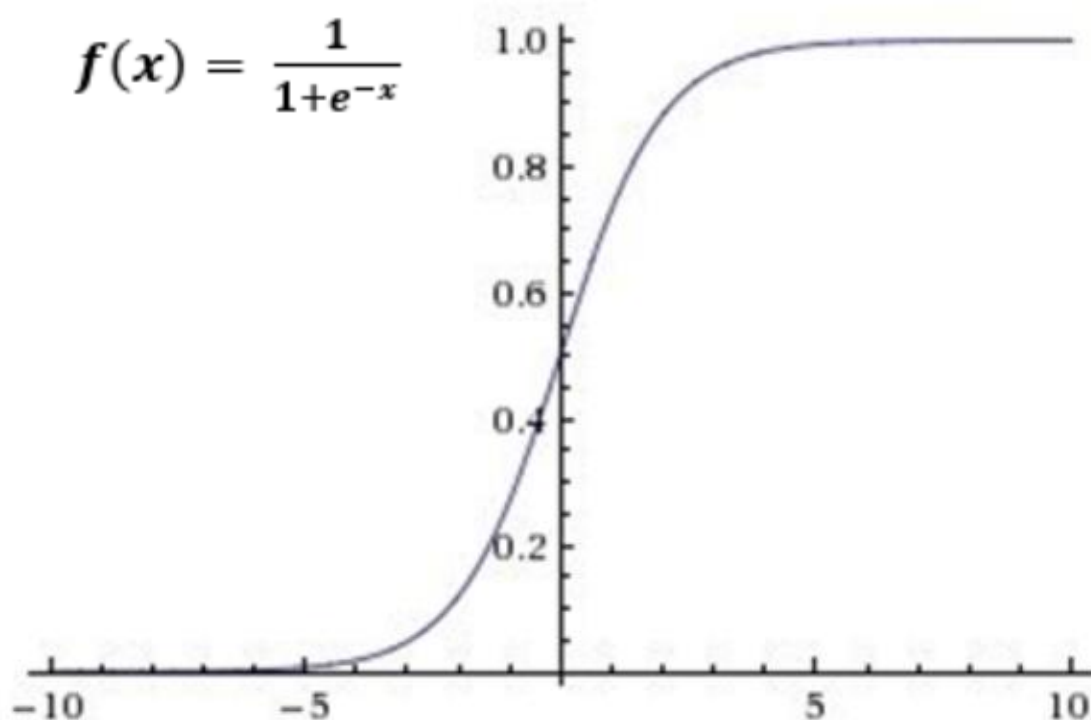


圖 3.6 線性整流函數 ReLU

輸出層(Output layer)為網路的最後一層，輸出層的神經元也可不止輸出一個，可以有多個輸出，這樣模型可以靈活的應用於分類回歸，以及其他的機器學習領域比如降維和聚類等，常見用於輸出層的激活函數包括 Softmax、Sigmoid。本研究使用 Softmax 作為輸出層激活函數，其輸出每個類別項目之機率會介於 0 到

1，輸出範圍有限，且優化穩定，求解容易，方程式如圖 3.7 所示。缺點為由於其軟飽和性，容易產生梯度消失，導致訓練出現問題，且輸出並不是以 0 為中心。

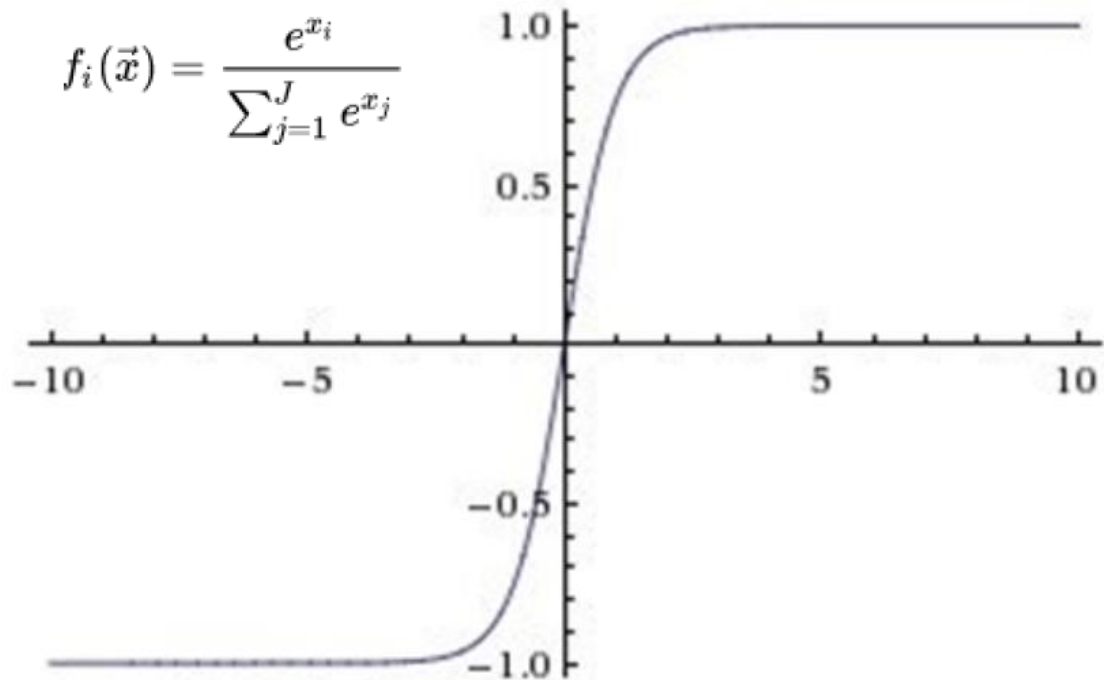


圖 3.7 Softmax 激活函數

3.2 情感分析

情感分析是分析人們對某些有興趣的實體，例如：產品、服務、組織、事件等的主觀看法(意見、情感、評價等)，並將這些資訊依據評論的對象與表達的主觀意見做摘要彙整，轉換成結構化的知識。

情感分析依其使用的技術主要可分為監督式與非監督式二種，監督式情感分析使用已標註之訓練資料來訓練分類器，藉由所訓練出的分類器進行目標資料的情感分類。非監督式情感分析則使用既有的情感辭典，或統計方法區別情感之正向、負向。情感識別為情感分析中一環，以往有關情緒分析相關研究，多將文本內容分成三元(正向、負向及中立)或二元(正向、負向)，三元情感分析有助於提高分類準確度，但仍無法探討出其情緒反應，如：開心、悲傷、驚訝、恐懼等。

從文獻回顧可得知，情感判別方法可分為兩大類：正負面情緒與 N 點量表。正負面情緒是將每則網路留言的意見傾向以正面、中性、或負面三個分類標記，並另外計算意見強度。而 N 點量表則是將每則網路留言的意見傾向以某個意見向度代表，數字同時代表意見強度，例如：正負面程度的分數，從-2、-1、0、+1 到+2 因此情感分析是一種「分類」的過程，每一個類別代表一種情緒反應，根據其規則將文本歸類為某種情緒反應。

本研究將使用 SnowNLP 套件作為情感分析的工具。SnowNLP 是一個 python 套件，可用來處理中文文本，主要功能包括：中文分詞、情感分析、提取關鍵詞、提取大綱等等。本研究將著重在使用情感分析的部分，套件內建的指令會給予文本 0 到 1 的分數，0 到 0.5 分表示負面，0.5 到 1 分表示正面。但由於套件內建語庫是針對買賣購物，因此本研究決定自訂語庫，透過爬蟲放入大量臺鐵相關的新聞、常見推文等作為鐵道版用語的訓練，以幫助更準確地分析留言的情感。

3.2.1 情感語料庫建立

首先分別建立正、負情感訓練用語料庫，圖 3.8 為正向情感語料庫，圖 3.9 為負向情感語料庫，內容皆為鐵道版常見推文及新聞，接著將兩個語料庫分別儲存回套件底下的根目錄進行訓練。

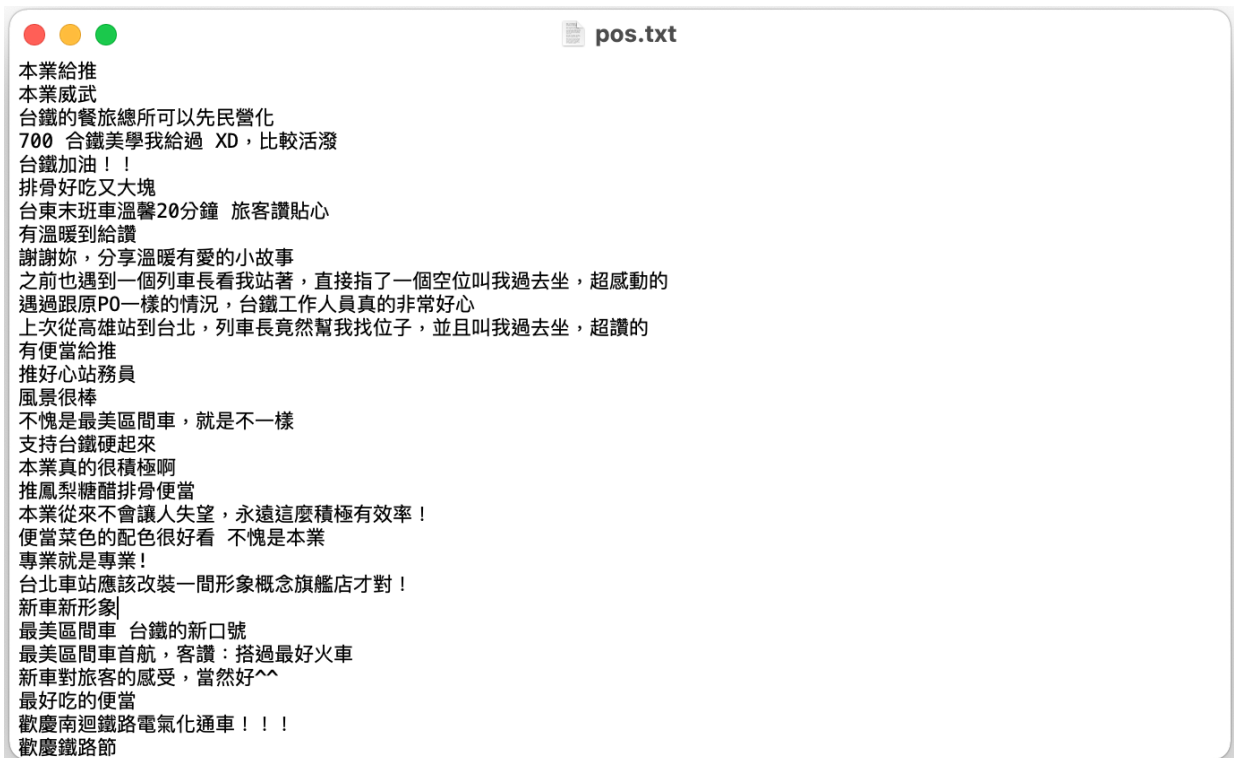


圖 3.8 正向情感語料庫



圖 3.9 負向情感語料庫

圖 3.10 為 SnowNLP 套件情感訓練的程式碼，包含分詞、詞性標註、情感分析過程，將訓練好的檔案儲存為 sentiment.marshall，之後修改套件裡的 data_path 指向訓練好的檔案即完成訓練。

```

1 # 中文分詞訓練庫
2 from snownlp import seg
3 seg.train('data.txt')
4 seg.save('seg.marshall')
5 #詞性標注訓練語料庫
6 from snownlp import tag
7 tag.train('train.txt')
8 tag.save('tag.marshall')
9 #情感分析訓練語料庫
10 from snownlp import sentiment
11 sentiment.train('neg.txt', 'pos.txt')
12 sentiment.save('sentiment.marshall')

```

圖 3.10 SnowNLP 套件情感訓練過程

表 3.1 為鐵道版推文例句，經由自訂語庫訓練前後的情感分數比較，可以發現下列例句大都為反諷語句，但原始語料庫計算出來的情感分數皆為正情感，而透過本研究自訂語料庫計算出來的情感分數皆變成負情感。

表 3.1 推文例句訓練前後情感分數比較

例句	原始分數	訓練後分數
珍惜生命遠離台鐵	0.9323602391	0.3258845225
台鐵不意外	0.8599315562	0.4122569855
每日任務達成	0.5807262312	0.3252591377
松鼠調皮了	0.8630530555	0.1190125322
連續假期 連續事故 台鐵真有你的	0.8411478078	0.2058925415
合鐵日常發揮而已	0.7925490769	0.3265495187
沒事沒事 便當有賣出去就好	0.7566235691	0.1854225607
便當局不是叫假的	0.7931515151	0.3731141392

3.2.2 權重設計

PTT 網站上每一篇貼文都有標題、本文、留言、推文數等架構，本研究將其分成兩大部分：「標題與本文」和「留言」。首先是標題與本文的部分，通常標題會與本文有關聯，由於本文的組成較為複雜，因此會先對本文截取一段摘要，並分別對本文的摘要以及標題進行情感分析，得到標題與本文的情感分數，此部分的權重為 0.5。

接著是留言的部分，會對每一則留言都個別進行情感分析，再根據留言推文顯示，推為+1，噓為-1(因為噓文的人很多是在反串)，箭頭為 0 來做為分數加總的正負號依據，得到留言的情感分數，而此部分的權重為 0.5。最後將兩部分分數加總後，即可得到該篇貼文的情感分析分數，然後再依推文數加權平均當天所有貼文的分數作為當天的總分，分數會落在 -1 到 1 之間，進而對當天評論進行情感分類。

3.2.3 情感辨識分類

以往研究多利用二元決策分類將情感值界定為正面與負面或採用三元決策分析，將情感值區分成正面、中立與負面。本研究為了增加情感分類細緻度，同陳翰(2016)將情感值設計為五個區間，情感辨識設計為：非常不滿意、不滿意、普通滿意、很滿意、非常滿意，各情感區間門檻值及顏色代表如圖 3.11 所示。

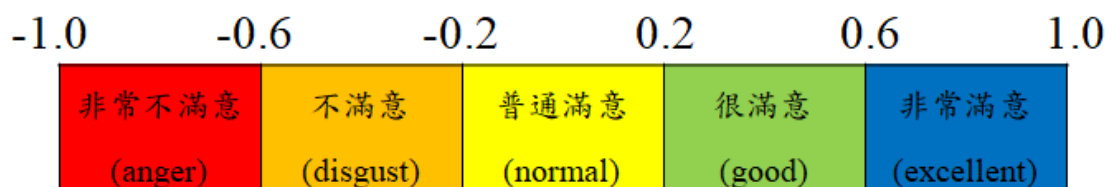


圖 3.11 情感區間尺度

3.3 安全績效與網路輿情關聯模型

根據文獻回顧，本研究使用皮爾森相關係數(Pearson's correlation coefficient)建立安全績效與網路輿情之關聯模型。相關係數很常用在機器學習或是統計分析上使用，主要衡量兩變數間「線性」關聯性的高低程度，探討兩個變數(或多變數)間是否存在線性關係。

皮爾森相關係數常用於探討兩連續變數(X, Y)之間的線性關係，其值介於-1到1之間，若兩變數之間的相關係數絕對值較大，則表示彼此相互共變的程度較大。一般而言，若兩變數之間為正相關，當X提升時，Y也會隨之提升；反之，若兩變數之間為負相關，則當X提升時，Y會隨之下降。皮爾森相關係數與共變異數公式如下所列，這邊的共變異數與變異數都是除上(n-1)而不是n，原因是只用少部分樣本在推論母體時因為偏量(bias)的關係，在推論時樣本推估會少一個自由度。

$$\text{相關係數(correlation coefficient)} : \rho = \frac{\text{x 和 y 的共變異數}}{\text{x 的標準差} \times \text{y 的標準差}}$$

$$\text{共變異數(covariance)} : \text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)$$

$$\text{變異數(variance)} : \text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2$$

$$\text{標準差(standard deviation)} : \text{std}(x) = \sqrt{\text{var}(x)}$$

皮爾森相關係數推導公式如下所列，相關係數(ρ ，有的時候會用r來表示)會落在-1到1之間： $-1 \leq \rho \leq 1$ ， ρ 為相關係數，COV為共變異數， μ_x 和 μ_y 分別代表變數x和y的平均數。

$$\begin{aligned}
\rho &= \frac{\text{cov}(x, y)}{\text{std}(x) \times \text{std}(y)} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\text{var}(x)} \sqrt{\text{var}(y)}} \\
&= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \mu_y)^2}} \\
&= \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} \\
&= \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sum_{i=1}^n (y_i - \mu_y)^2}
\end{aligned}$$

從公式來看，將 x 和 y 減去各自的平均數後相乘，最後算出總和，如果假設變數 x 等於變數 y 時，此時共變異數就等於變異數，這時候 ρ 的值就等 1，所以 x 和 y 就完全相關($x=y$)。

皮爾森相關係數的變化範圍為-1 到 1 之間。係數的值為 1 意味著 X 和 Y 可以很好的由直線方程式來描述，所有的數據點都很好的落在一條直線上，且 Y 隨著 X 的增加而增加。係數的值為-1 表示所有的數據點都落在直線上，且 Y 隨著 X 的增加而減少。另外，係數為 0 代表兩個變數之間沒有線性關係。

若 X_i 和 Y_i 均落在他們各自均值的同一側，則 $(X_i - X)(Y_i - Y)$ 的值為正。也就是說，如果 X_i 和 Y_i 同時趨向於大於，或同時趨向於小於他們各自的均值，則相關係數為正；如果 X_i 和 Y_i 趨向於落在他們均值的相反一側，則相關係數為負。本研究使用 SAS 軟體進行相關係數計算。

第四章 實證分析

本研究將藉由實證分析，進行民眾對臺鐵服務之評論意見挖掘，建立結合安全績效與網路輿情之關聯模式，並以視覺化儀表板呈現。首先從建立以臺鐵為本體的社群媒體評論資料庫開始，包括留言情感分類、內容分類，最後將意見挖掘成果以視覺化展現，以此建立一套有效的社群媒體意見挖掘系統，用於探索軌臺鐵服務評論的意涵，供臺鐵管理階層易於了解使用者的意見。

實證分析流程如圖 4.1 所示。資料收集分為安全績效、網路輿情兩部分，安全績效資料來自臺鐵提供的行車事故報表資料，本研究針對事故類型、傷亡人數進行統計；網路輿情資料來自 PTT Railway 版貼文，經過文本處理、情感分析後與安全績效資料進行比較。最後合併成關聯模型、製作視覺化儀表板。

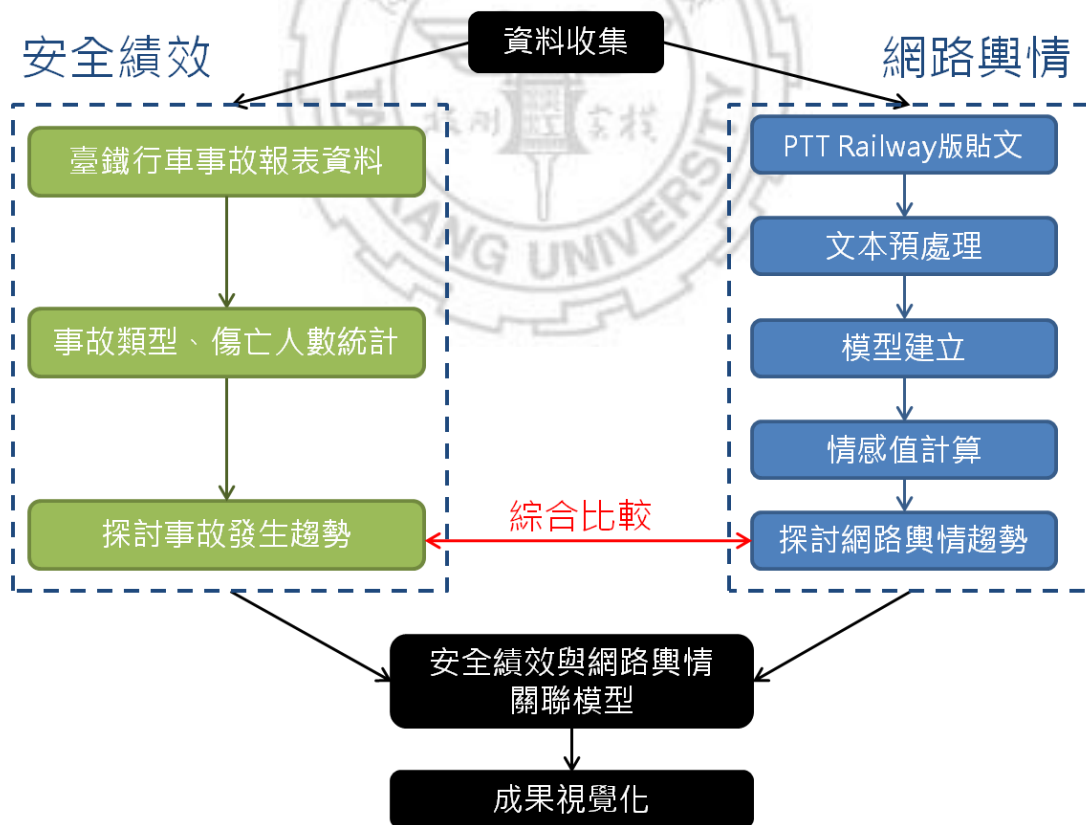


圖 4.1 實證分析流程圖

4.1 以臺鐵為本體的社群媒體資料庫建立

本研究透過自編程式網絡爬蟲蒐集 Ptt 實業坊 Railway 版上有關臺鐵的相關貼文，資料蒐集時間訂為 2018 年 1 月 1 日至 2021 年 5 月 7 日止，總共蒐集 2,054 篇貼文，84,779 則留言，並將評論主題分為「車站設施」、「員工權益」、「餐飲服務」、「票證系統」、「列車運轉」五大類相關文本。

4.1.1 爬蟲系統設計

本研究根據人工設立之關鍵字進行計程收費網路輿情評論文本爬蟲系統的設計，依據抓回之文本結合人工於來源網站抽驗檢視結果，發現文本數量短缺時即修正關鍵字組合，依據以下方式進行爬蟲關鍵辭庫之設計：

- (1) 車站設施：與臺鐵車站硬體設施相關議題與名詞的組成
- (2) 員工權益：與臺鐵員工工作內容、福利、改革等相關議題與名詞的組成
- (3) 餐飲服務：與臺鐵便當相關議題與名詞的組成
- (4) 票證系統：與臺鐵票務相關議題與名詞的組成
- (5) 列車運轉：與臺鐵列車運轉及軌道線路相關議題與名詞的組成

採用人工定義關鍵詞的方法，將前述五類項目建立關鍵辭庫，並加入聯合規則以提升爬蟲搜尋文章知精確度，關鍵詞庫如表 4.1 所示。

表 4.1 爬蟲關鍵字詞庫

關鍵詞庫	關鍵詞設定
車站設施	車站 & [台鐵、臺鐵]的交集
員工權益	[員工/福利/人力/工時/司機/車長/站務/薪水/津貼] & [台鐵、臺鐵]的交集
餐飲服務	[便當/餐飲] & [台鐵、臺鐵]的交集
票證系統	[訂票/售票/電子票/站票/退票/換票/紀念票] & [台鐵、臺鐵]的交集
列車運轉	[列車/車廂/異常/故障/事故/區間車/自強/莒光/復興/太魯閣/普悠瑪/EMU] & [台鐵、臺鐵]的交集

透過上述關鍵詞對爬蟲系統抓回之文章進行過濾，依前述章節所提到之分類項建立初步分類文本資料庫，再透過人工檢視刪除不相關文本，並給予相關文本簡短結論後，再存回分類資料庫，如此爬蟲分類資料庫下之文本即不需要再進行文本去噪工作，可減少在文本預處理時因去噪方法的不同所造成的分析問題。

4.1.2 資料內容說明

本研究自 2018 年 1 月 1 日至 2021 年 5 月 7 日止，針對 Ptt 實業坊 Railway 版上有關臺鐵的相關貼文抓取文本資料，共 2,054 篇貼文，包含：列車運轉類 927 篇、餐飲服務類 120 篇、員工權益類 266 篇、車站設施類 326 篇、票證系統類 415 篇，如圖 4.2 所示。留言共收集 84,779 則，包含：列車運轉類 40,000 則、餐飲服務類 3,639 則、員工權益類 13,669 則、車站設施類 11,624 則、票證系統類 15,847 則，如圖 4.3 所示。

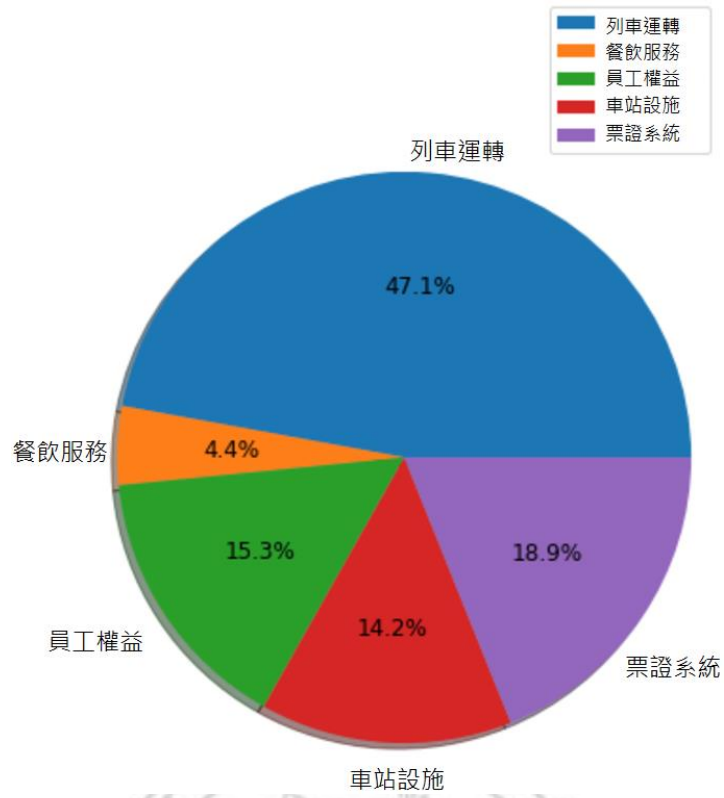


圖 4.2 收集到的貼文總數占比

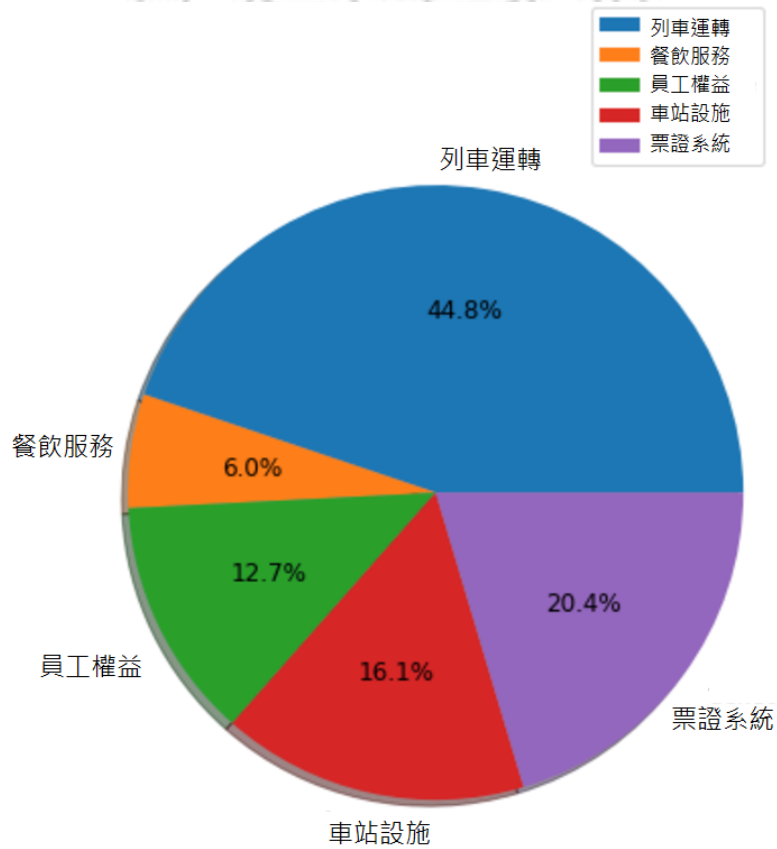


圖 4.3 收集到的留言總數占比

相較於陳亭愷(2015)、張悅朗(2020)兩者研究，本研究之文本內容經過人工快速檢視，可節省大量時間。圖 4.4 為網路爬蟲蒐集回來的臺鐵輿情文本列表範例，共有日期、推文數、標題、網址連結、內文等五個欄位。

date	push	title	link	content
120	2/14	23	[閒聊] 台鐵台中車站站外天橋廣場太髒 https://www.ptt.cc/bbs/Railway/M.1613262367.A...	如題，初三回台中過年，出了台中站，到了外面，天橋平台廣場建設景觀很棒，但地面到處都是...
121	2/13	9	[新聞] 台鐵422次神秘列車台北到潮州擁有無敵敵 https://www.ptt.cc/bbs/Railway/M.1613187654.A...	台鐵422次神秘列車 台北到潮州擁有無敵海景（中央社記者盧太城台東縣13日電）南...
122	2/12	3	[問題] 台鐵電子車票多取一張票 https://www.ptt.cc/bbs/Railway/M.1613143884.A...	如題 剛剛買了兩張票（分開買）自己已經取電子車票 原本另一張是要分給家人...
123	2/12	3	[閒聊] 台鐵 媽祖彩繪列車 https://www.ptt.cc/bbs/Railway/M.1613127724.A...	有圖有真相網誌版： https://osakaleo.pixnet.net/blo...
124	2/12	9	[新聞] 台鐵戰士 / 六姊妹扛粗活泥淖裡不讓鬚眉 https://www.ptt.cc/bbs/Railway/M.1613108545.A...	台鐵戰士 / 六姊妹扛粗活 泥淖裡不讓鬚眉 https://udn.com/news/...
125	2/11	3	[新聞] 四腳亭=瑞芳車站班次延誤 台鐵：自主檢 https://www.ptt.cc/bbs/Railway/M.1613026226.A...	https://udn.com/news/story/7266/5247589.2...
126	2/11	12	Re: [閒聊] 私心建議 覺得台鐵這樣就不會嚴重誤點了 https://www.ptt.cc/bbs/Railway/M.1612984996.A...	*引述《yuni384257 (Edge)》之銘言：：標題: [閒聊] 私心...
127	2/09	10	[新聞] 台鐵ATP採購案 瑞典承包商提告獲賠 https://www.ptt.cc/bbs/Railway/M.1612855968.A...	台鐵ATP採購案 瑞典承包商提告獲賠 https://news.ltn.com...

圖 4.4 臺鐵輿情文本列表範例

4.2 安全績效資料

本研究選取的安全績效資料是指臺鐵提供的行車事故報表資料。首先將每日發生的事故與事件數做統計並畫成直方圖，如圖 4.5 所示，藍色線為行車異常事件數，紅色線為重大行車事故及一般行車事故數。接著統計每次事故與事件的傷亡人數，如圖 4.6 所示，紅色為死亡人數，藍色為受傷人數，從圖中可發現除 2021 年 4 月 2 日發生的太魯閣號出軌事故造成重大傷亡外，其他事故大都只有零星傷亡人數，且造成死亡的原因都是發生平交道事故。

安全績效

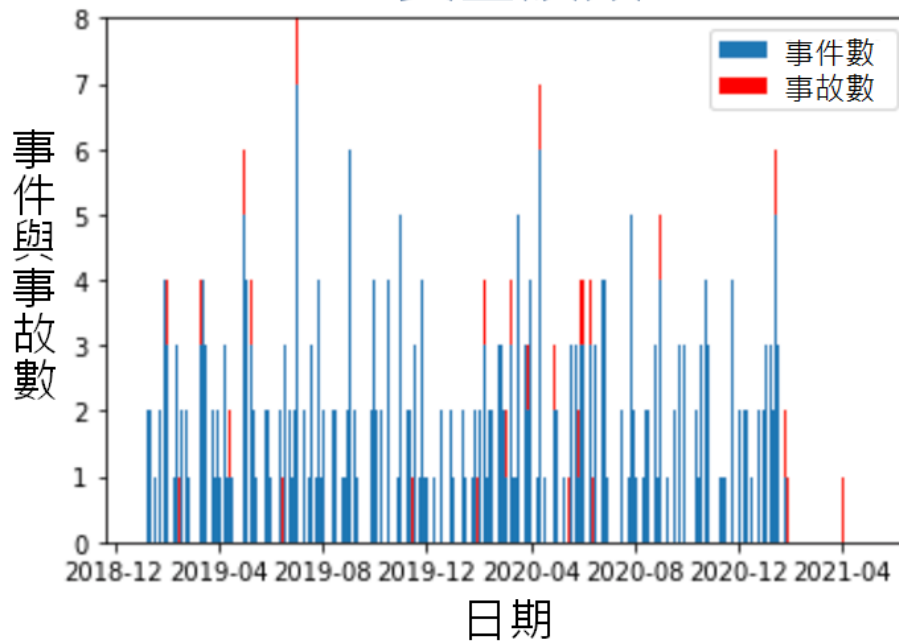


圖 4.5 行車異常事件與行車事故統計

傷亡人數

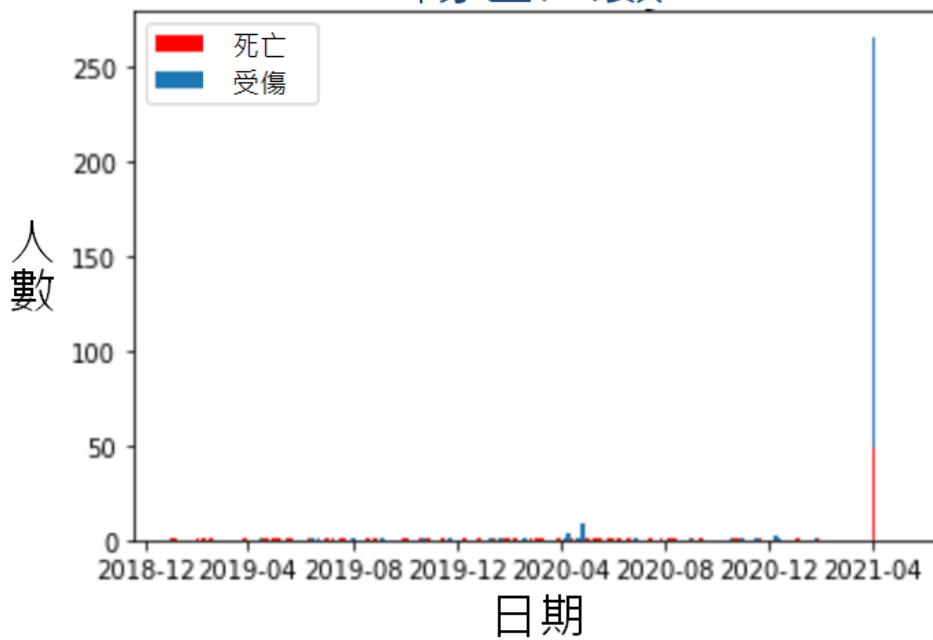


圖 4.6 事故傷亡人數統計

4.3 網路輿情資料

網路輿情資料是由網絡爬蟲蒐集 Ptt 實業坊 Railway 版上有關臺鐵的相關貼文，本節將五大類留言分別探討如下：

4.3.1 車站設施類別

圖 4.7 為車站設施類別的留言數，該圖顯示 2020 年 2 月 27 日的留言數最高，共有 301 則留言，原因是該日鐵道局宣布基隆輕軌無法採用 Tram-Train 系統，迫使基隆車站必須要輕軌、臺鐵二選一，網友便熱烈討論臺鐵是否將捨棄基隆車站的話題，有些民眾認為基隆車站還是要由臺鐵來服務，方便基隆民眾前往西部各縣市，但部分民眾認為臺鐵服務品質不佳且時常誤點，營運班次亦不如輕軌密集，離峰時段候車時間至少半小時，希望由輕軌取代臺鐵，因此引來兩方在網路上激烈辯論。

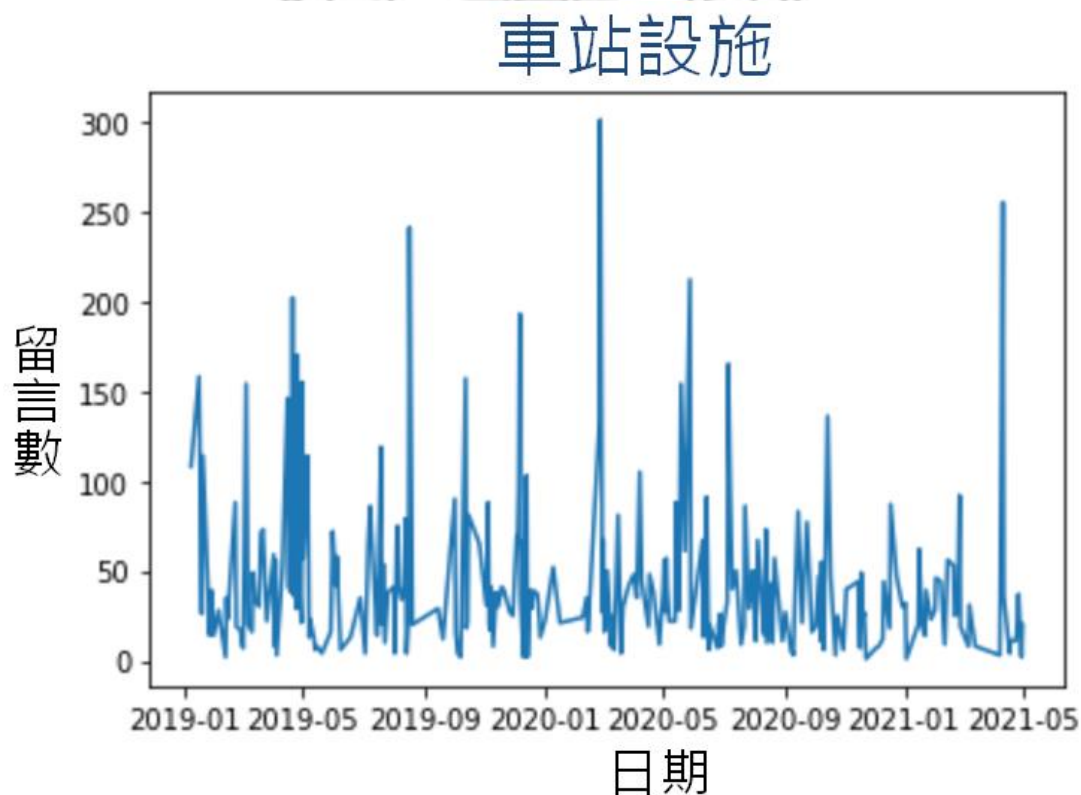


圖 4.7 車站設施類別留言數

圖 4.8 為車站設施類別的留言情感分數，該圖顯示同樣在 2020 年 2 月 27 日的留言數情感分數為兩年來最低，情感分數為-133 分，顯示民眾對於基隆車站必須要輕軌、臺鐵二選一的決策不盡滿意，多數民眾仍希望輕軌計畫不應影響臺鐵現有的營運，且基隆新站才剛完工啟用，若要拆除改建，無疑是浪費公帑。

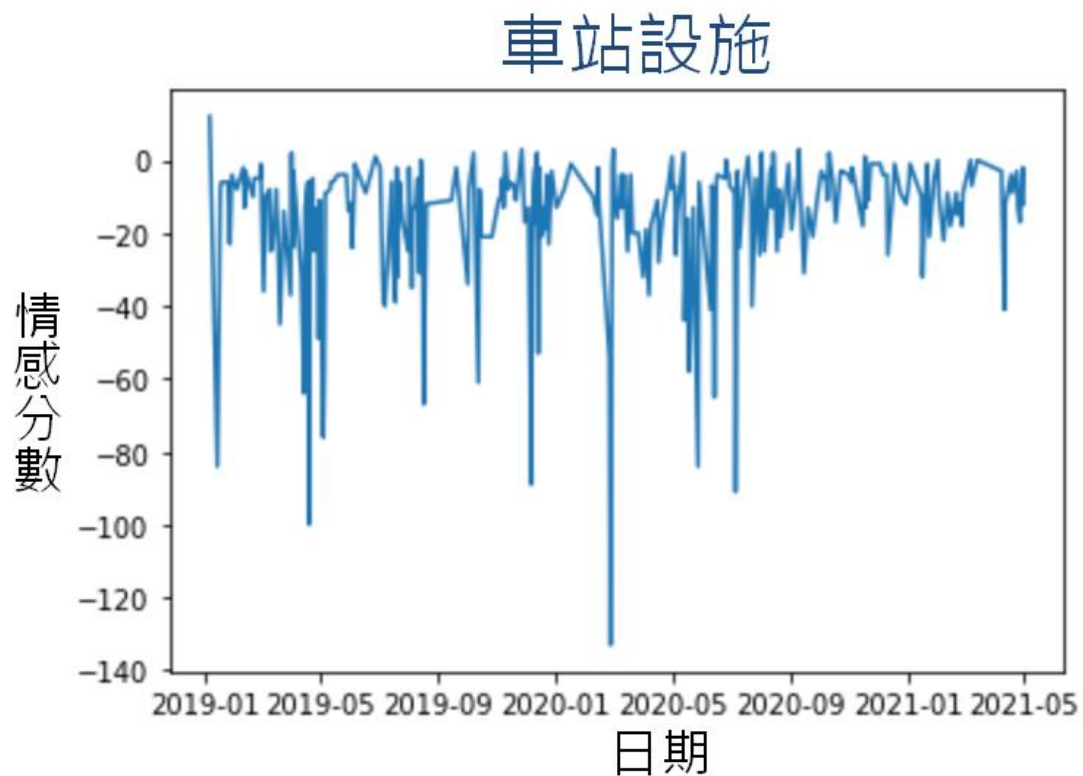


圖 4.8 車站設施類別情感分數

4.3.2 員工權益類別

圖 4.9 為員工權益類別的留言數，該圖顯示 2021 年 3 月 16 日的留言數為最多，共有 451 則留言，此係因是該日臺鐵工會宣布要在清明假期發起大罷工，抗議臺鐵要推動四輪三班制，落實周休二日。但此制度可能讓員工月薪減少一萬一千元，因此臺鐵企業工會宜蘭分會發起連署，擬於清明連假依法休假，預計有超過二百人響應，將衝擊東部清明連假交通疏運。對此，交通部強調輪班制度在團體協約有規範，明年之前不會變動，請部分工會員工能夠配合疏運，網路上亦有不少民眾持正反意見，有人支持工會行動，也有民眾認為不應影響連假疏運。

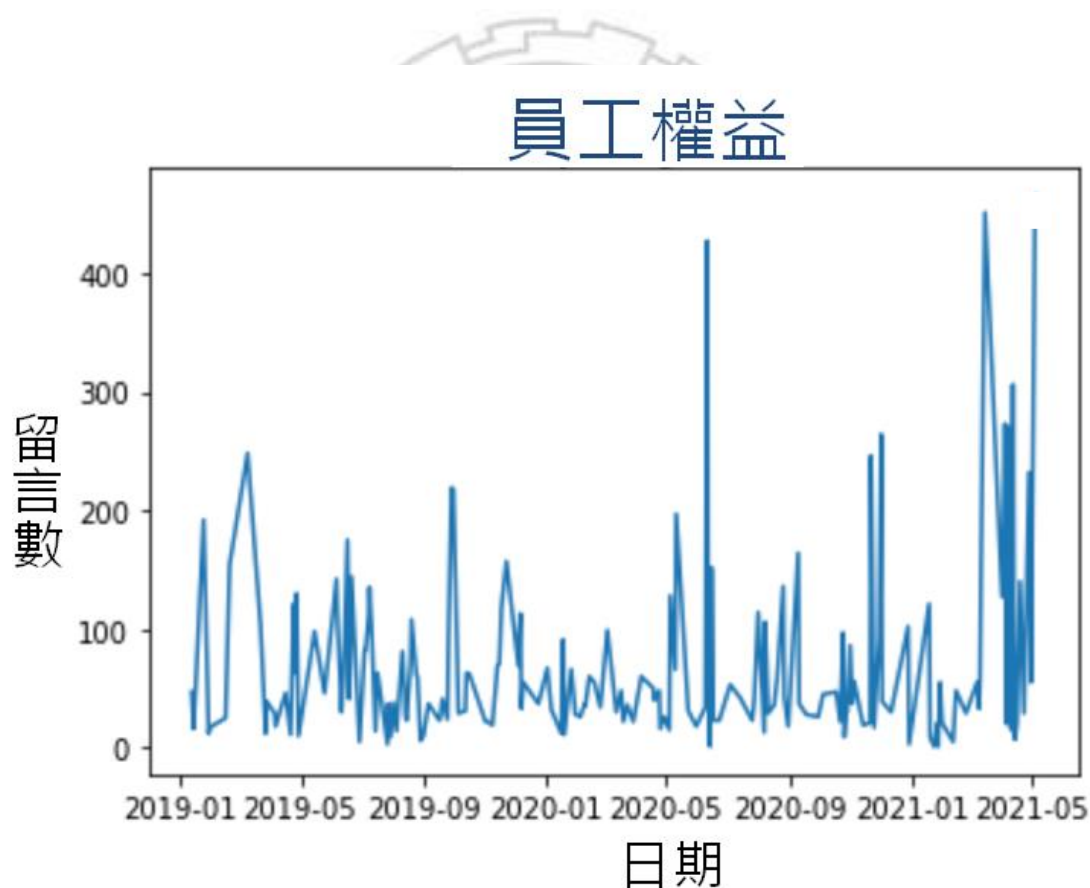


圖 4.9 員工權益類別留言數

圖 4.10 為員工權益類別的留言情感分數，該圖顯示在 2020 年 6 月 11 日的留言數情感分數為兩年來最低，情感分數為-237 分，原因是該日有民眾到嘉義火車站買票，並詢問相關問題，未料售票員語氣甚為不友善，讓民眾覺得莫名其妙，認為售票員態度不佳，也引發下方留言網友撻伐。

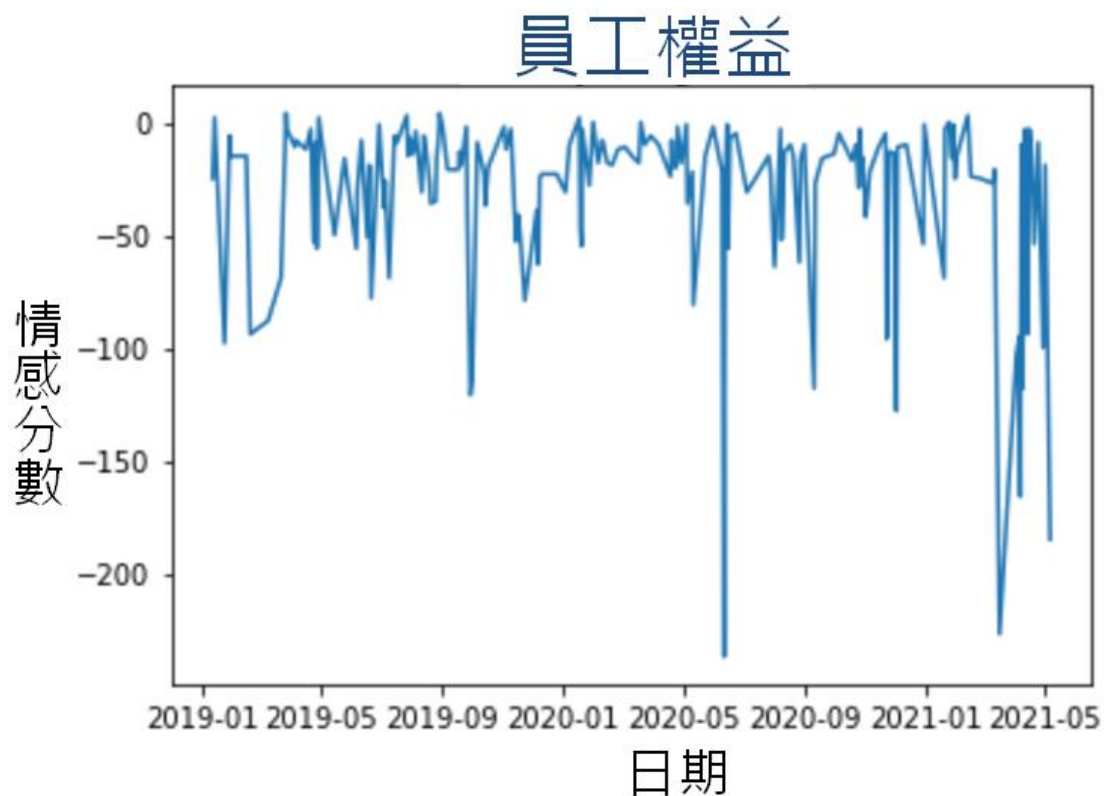


圖 4.10 員工權益類別留言情感分數

4.3.3 餐飲服務類別

圖 4.11 為餐飲服務類別的留言數，該圖顯示 2020 年 9 月 23 日的留言數來到最高，共有 125 則留言，原因是臺鐵為自家的鐵路便當推出全新的品牌視覺識別系統與行銷策略，盼能提升臺鐵便當本舖整體品牌形象好感度，並透過整合統一行銷廣告宣傳，加深旅客與消費者正面印象。臺鐵便當每年賣破千萬個，為臺鐵進帳七億元，被視為臺鐵副業的「金雞母」。此一新聞立刻在網路上引起廣大的回響，臺鐵每年賣出上千萬個便當，銷售量亮眼，確實是一個重要的副業，留言多支持「台鐵的品牌價值展現在便當上」。

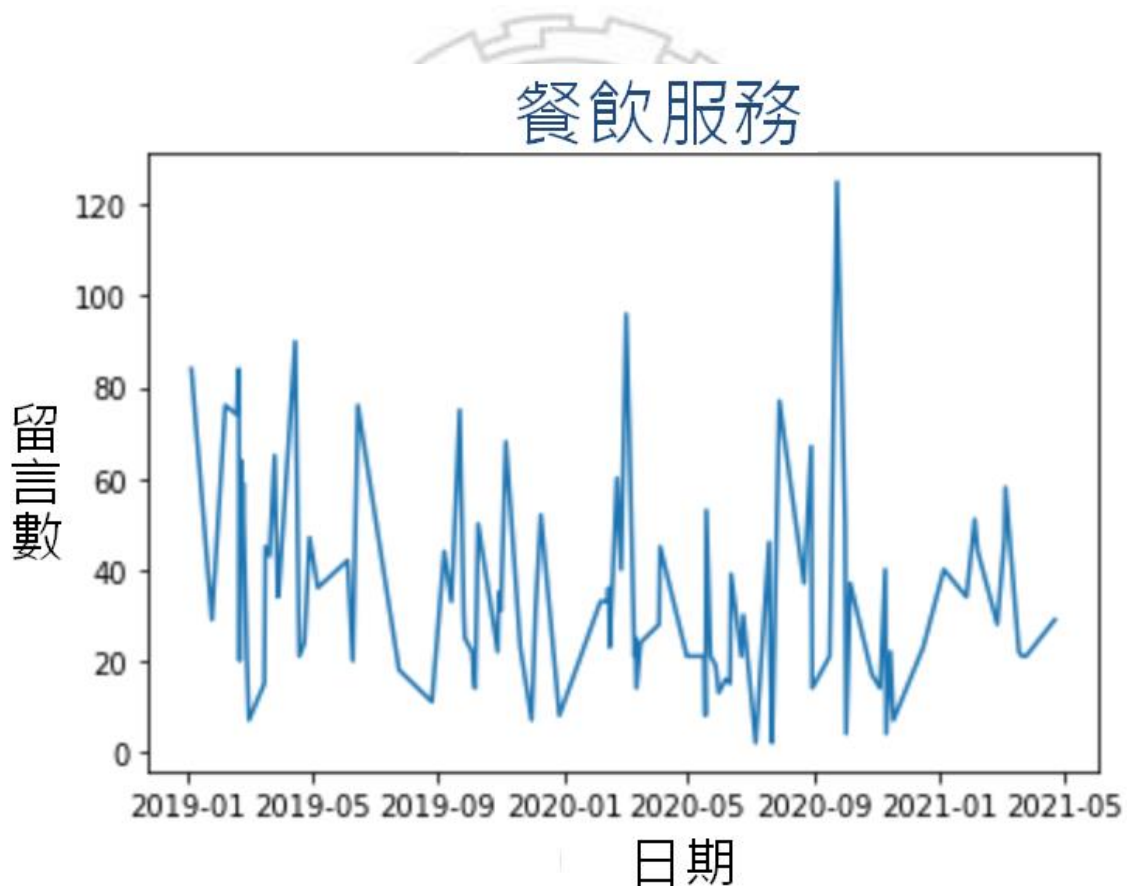


圖 4.11 餐飲服務類別留言數

圖 4.12 為餐飲服務類別的留言情感分數，該圖顯示相較於其他四類，餐飲服務類別的留言具有較高的正向情感分數，顯示臺鐵在餐飲服務方面的表現較博得民眾青睞，原因來自於臺鐵便當的名氣。圖中唯一最低的負評發生在 2020 年 3 月 2 日，因為該日有民眾留言在車站購買臺鐵便當無法使用行動支付，引來許多網友在下方留言，認為臺鐵服務太落伍，公營事業跟不上時代的腳步。

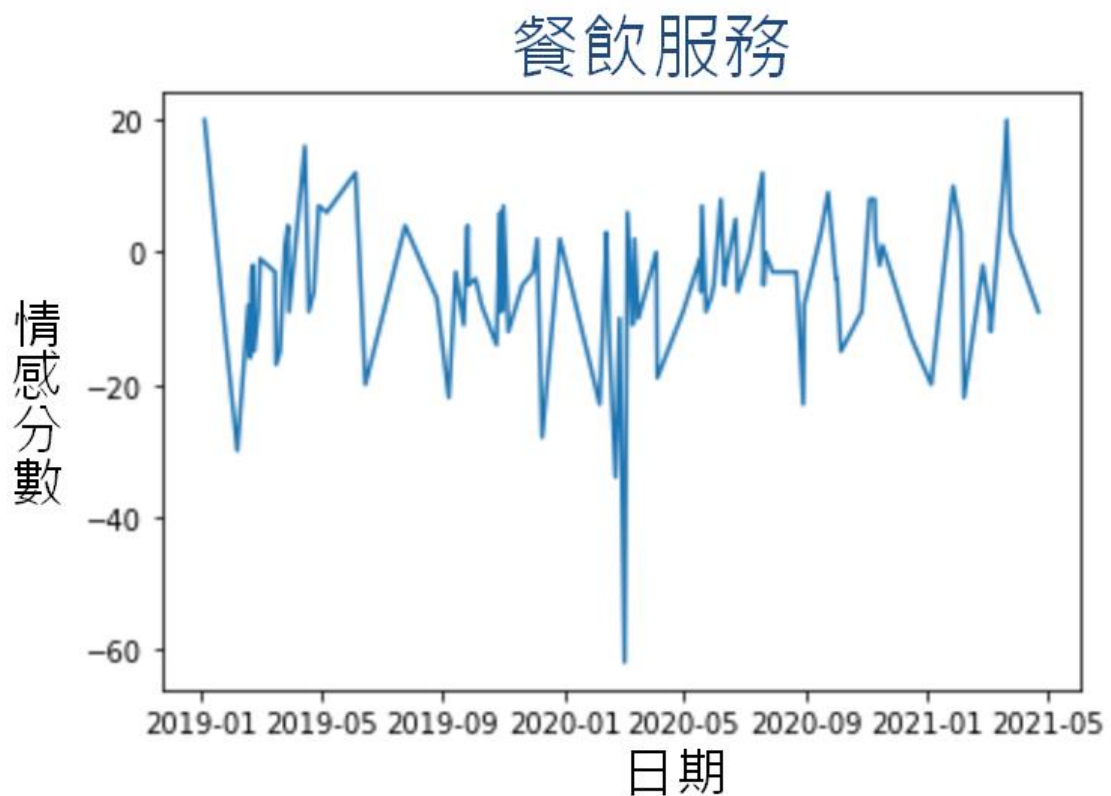


圖 4.12 餐飲服務類別留言情感分數

4.3.4 票證系統類別

圖 4.13 為票證系統類別的留言數，該圖顯示 2019 年 4 月 29 日的留言數來到最高，共有 391 則留言，原因是臺鐵為紓解旅客乘車之需求，將於 2019 年 5 月 2 日起開放全線普悠瑪號、太魯閣號等新自強號於乘車當日每列次限量發售 120 張無座票，此舉引發民眾質疑，難道臺鐵之前說為了提升服務品質以及安全因素不開放站票均未屬實，在網路上掀起一陣熱議。留言數次高的日期是 2019 年 12 月 15 日，當日因民眾質疑臺鐵超賣站票導致一群有事前購買座票旅客無法上車，且發生問題第一時間站方處理態度非常消極，需權益受損旅客自行出站進行排隊與服務中心聯繫，引起民眾相當不滿，表示臺鐵不論是事前的危機預防還是事情發生後的應對處理方式，都引來批評。從上述兩例顯示臺鐵的站票問題，時常引發民眾的不滿，臺鐵確實需要檢討其站票規定是否合宜。

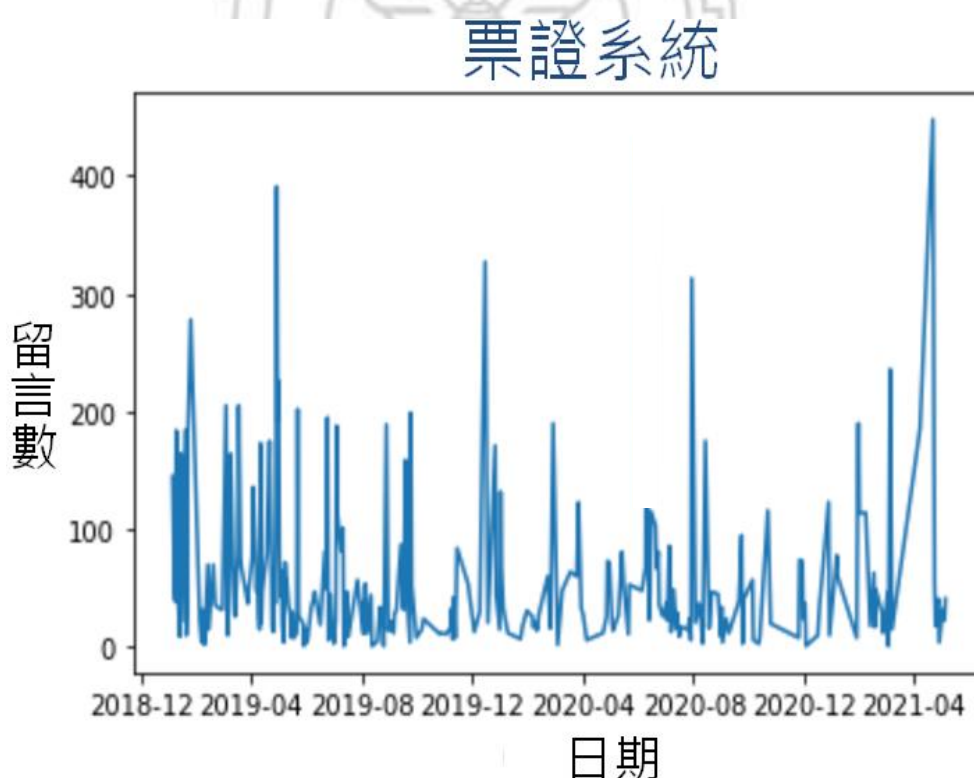


圖 4.13 票證系統類別留言數

圖 4.14 為票證系統類別的留言情感分數，該圖顯示在 2020 年 7 月 30 日的留言情感分數為兩年來最低，情感分數為-209 分，原因是當天有兩篇關於換票、一篇關於補票的新聞。該日有民眾提早到車站想要換比較早的班次，結果因為該班車已沒有座位，售票員好心幫他換成三段車票，卻被乘客質疑明明是「換票」，為何要多收「退票」手續費，讓民眾質疑這樣的退票手續費規定是否合理？對此臺鐵表示退票手續費會提高至兩成，主要也是因為先前多次發生有人當天才退票，使得普悠瑪跟太魯閣號空位一大堆造成損失，因此臺鐵祭出新的退票規定。另一則新聞則是先前一位民眾發現自己搭錯車，卻因為錢不夠無法補票，與列車長發生爭執，因此挨告違反社會秩序維護法，不過，台中地院法官認為，整起事件的起因是台鐵車次誤點，導致該旅客搭錯車，最終裁定不罰。上述兩起事件皆引起網友熱議，使得該日留言情感分數為兩年來最低。

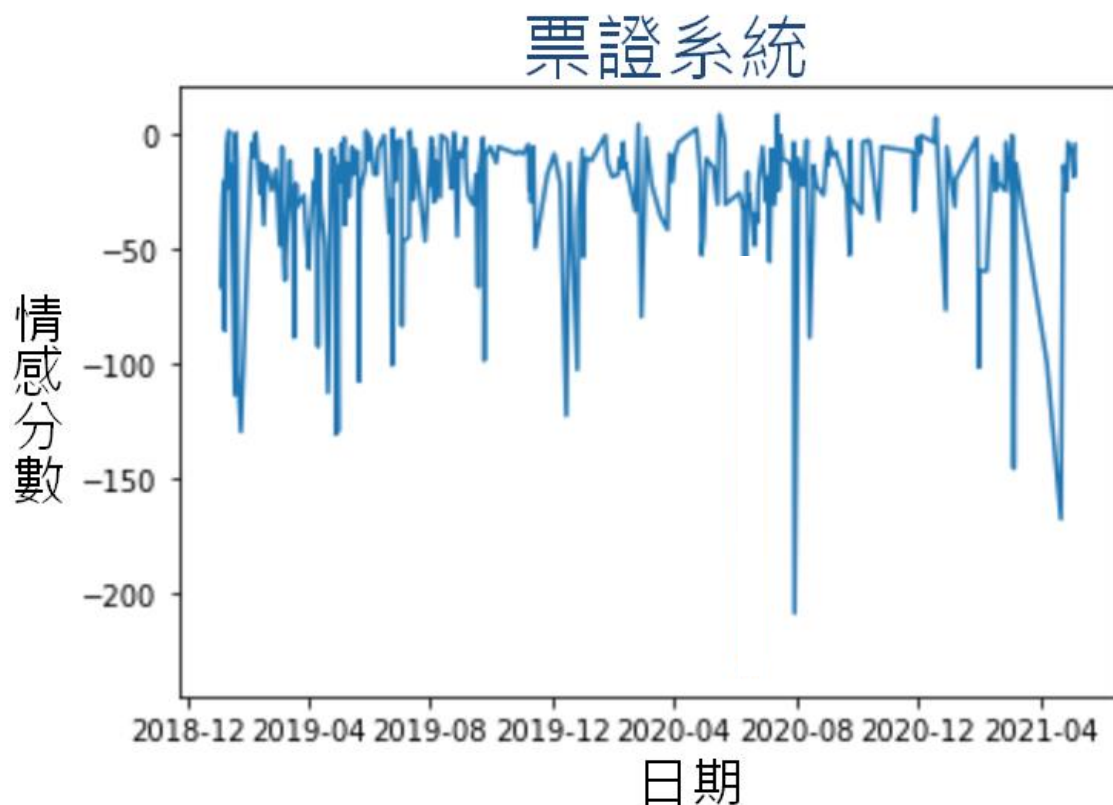


圖 4.14 票證系統類別留言情感分數

4.3.5 列車運轉類別

圖 4.15 為列車運轉類別的留言數。2021 年 4 月 2 日 408 次太魯閣號列車在行經花蓮大清水隧道時，被一輛停車不慎沿著鐵軌上方邊坡滑落的工程車撞上，造成列車失控出軌擦撞山壁，多節車廂扭曲變形，釀 49 死、200 多傷的嚴重傷亡，這也是臺鐵史上死傷最多的事故。事故發生後隔日 4 月 3 日的留言數來到 3,184 則，為兩年來最高，但到 4 月 7 日討論熱度就逐漸下滑，只剩 363 則留言。從圖中可看出當重大事故發生時網路輿情的飆漲非常迅速，但衰退速度亦甚快。

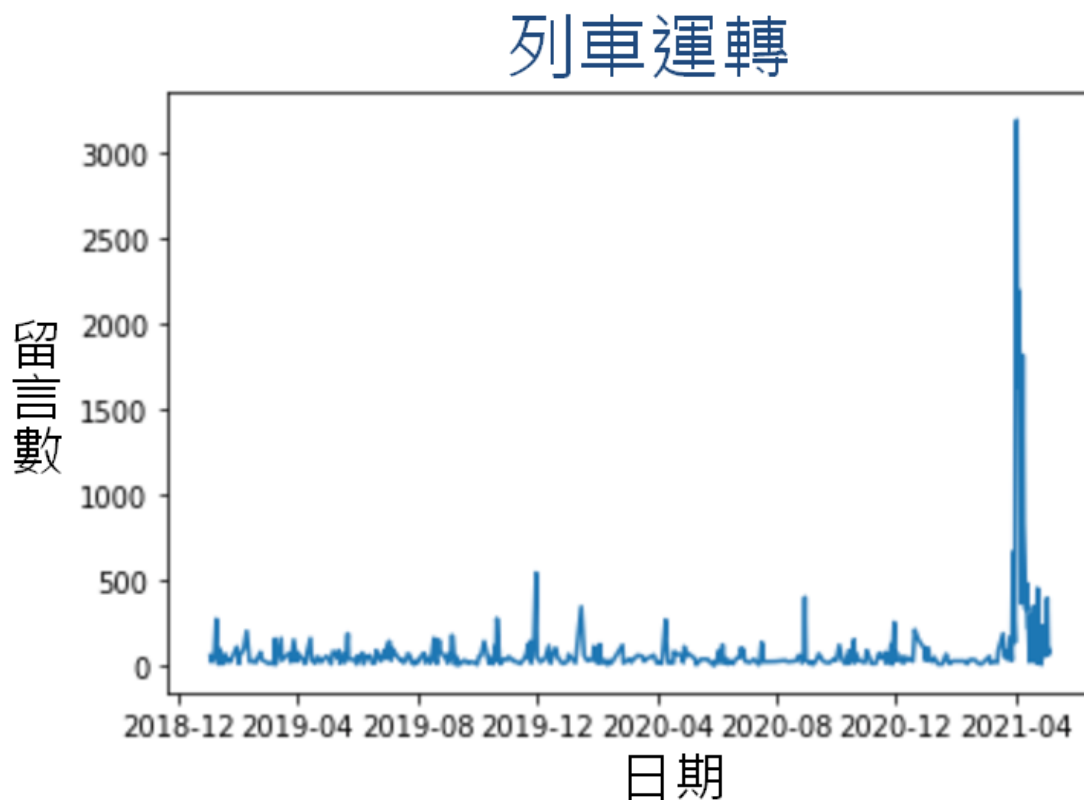


圖 4.15 列車運轉類別留言數

圖 4.16 為列車運轉類別的留言情感分數，該圖顯示太魯閣號事故發生後讓國人悲憤不已，從網路留言的情感趨勢可看出明顯變化，在 2021 年 4 月 3 日的留言數情感分數為兩年來最低，情感分數為-2554 分，為兩年來最低，其他日期的情感分數皆落在 10 到-200 分之間。

由圖 4.15 及圖 4.16 可發現情感分數與留言數剛好成反比趨勢，留言數越多情感分數就越低，符合網路輿情貶多於褒之特性。

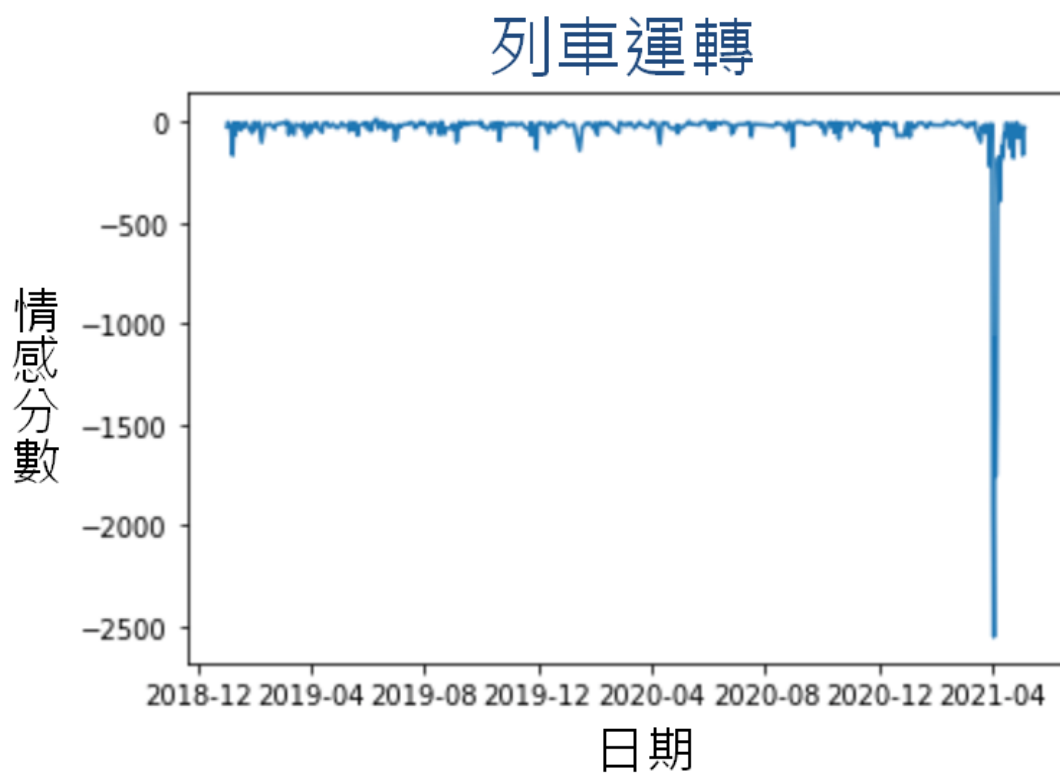


圖 4.16 列車運轉類別留言情感分數

4.3.6 網路輿情結果分析

網路輿情趨勢分析方法可用於臺鐵輿情文本分析，透過分析民眾情感的傾向程度、關注話題是否隨時間的推移而成長或消散、話題下的輿情情感傾向是否轉向，可了解民眾對臺鐵服務的滿意程度。

圖 4.17 為網路輿情五大評論類別正規化後的情感分數分布圖，分數落在-1到 1 之間，該圖顯示五大類別的評論皆落在 0 到-1 之間比較多，表示負面評論遠多於正面評論，符合網路輿情貶多於褒之特性。

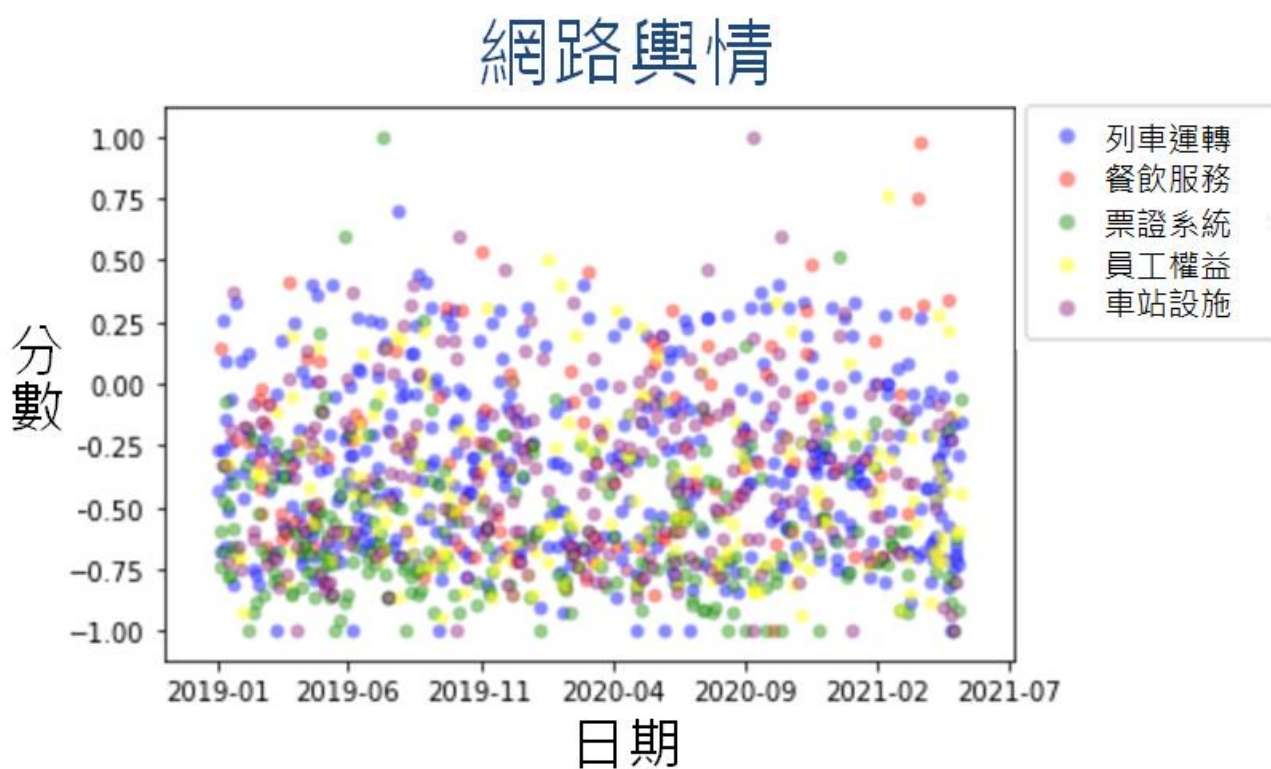


圖 4.17 五大評論類別情感分數分布圖

圖 4.18 為每日留言情感占比，本研究將每日評論情感分類為五個類型，包括：非常不滿意(anger)、不滿意(disgust)、普通滿意(normal)、很滿意(good)、非常滿意(excellent)。其中情感傾向以非常不滿意型數量最多高達 529 次，非常滿意型最少只有 8 次，反映整體旅客對臺鐵服務的情感傾向非常不滿意。

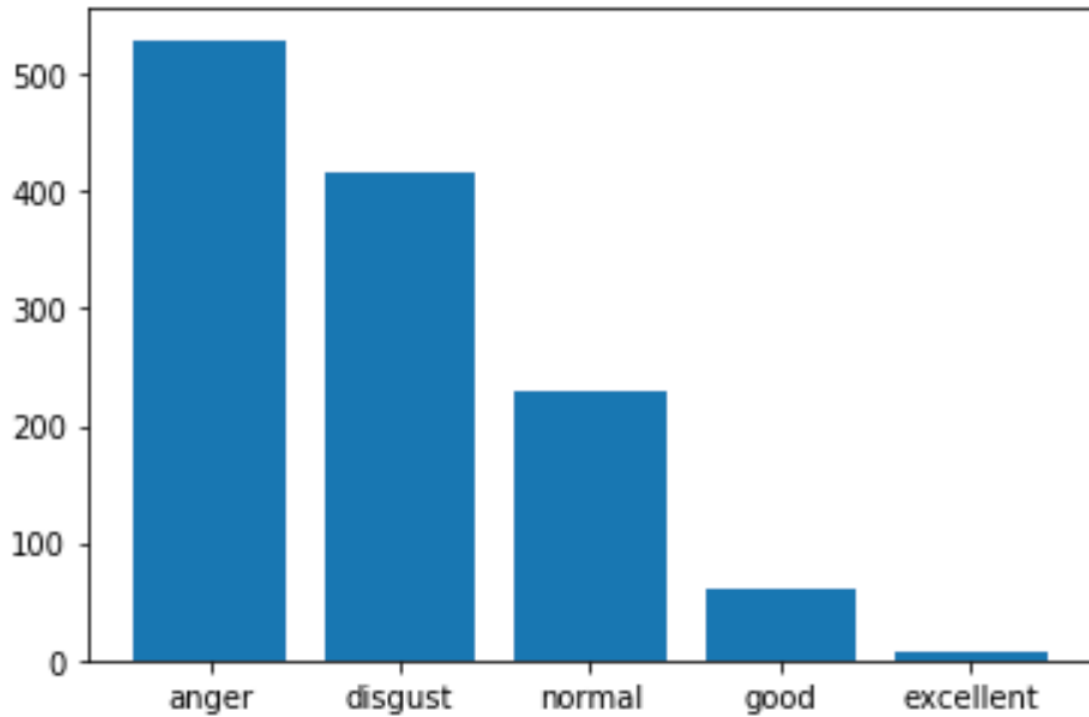


圖 4.18 留言情感占比

4.4 安全績效與網路輿情關聯模型

安全績效與網路輿情關聯模型是由 SAS 統計分析軟體進行相關分析，探討安全績效與網路輿情之間的關聯性，本節將網路輿情五大類留言分別探討如下：

4.4.1 車站設施類別

如圖 4.19 所示，車站設施類別的情感分數與總事件數無相關性，P 值為 0.2131，大於 0.0001 代表不具顯著性，表示事故或事件的發生不會影響車站設施類別的情感分數。



圖 4.19 車站設施類別情感分數與總事件數相關分析

4.4.2 員工權益類別

如圖 4.20 所示，員工權益類別的情感分數與總事件數無相關性，P 值為 0.6798，大於 0.0001 代表不具顯著性，表示事故或事件的發生不會影響員工權益類別的情感分數。



圖 4.20 員工權益類別情感分數與總事件數相關分析

4.4.3 餐飲服務類別

如圖 4.21 所示，餐飲服務類別的情感分數與總事件數無相關性，P 值為 0.2629，大於 0.0001 代表不具顯著性，表示事故或事件的發生不會影響餐飲服務類別的情感分數。



圖 4.21 餐飲服務類別情感分數與總事件數相關分析

4.4.4 票證系統類別

如圖 4.22 所示，票證系統類別的情感分數與總事件數無相關性，P 值為 0.2138，大於 0.0001 代表不具顯著性，表示事故或事件的發生不會影響票證系統類別的情感分數。



圖 4.22 票證系統類別情感分數與總事件數相關分析

4.4.5 列車運轉類別

如圖 4.23 所示，列車運轉類別的情感分數與總事件數呈高度負相關，相關係數 r 值為 -0.68669 ， P 值小於 0.0001 代表具有顯著性，表示事故或事件的發生會影響列車運轉類別的情感分數，發生事故越多，情感分數就越低。



圖 4.23 列車運轉類別情感分數與總事件數相關分析

4.4.6 安全績效與網路輿情關聯模型小結

如圖 4.24 所示，留言情感總分與總留言數呈高度負相關，相關係數 r 值為 -0.98445 ， P 值小於 0.0001 代表具有顯著性，表示留言數量多寡會影響留言情感總分，且五大類別評論皆是，留言數越多，情感分數就越低，符合網路輿情貶多於褒之特性。



圖 4.24 留言情感總分與總留言數相關分析

4.5 安全績效與網路輿情之視覺化對照圖

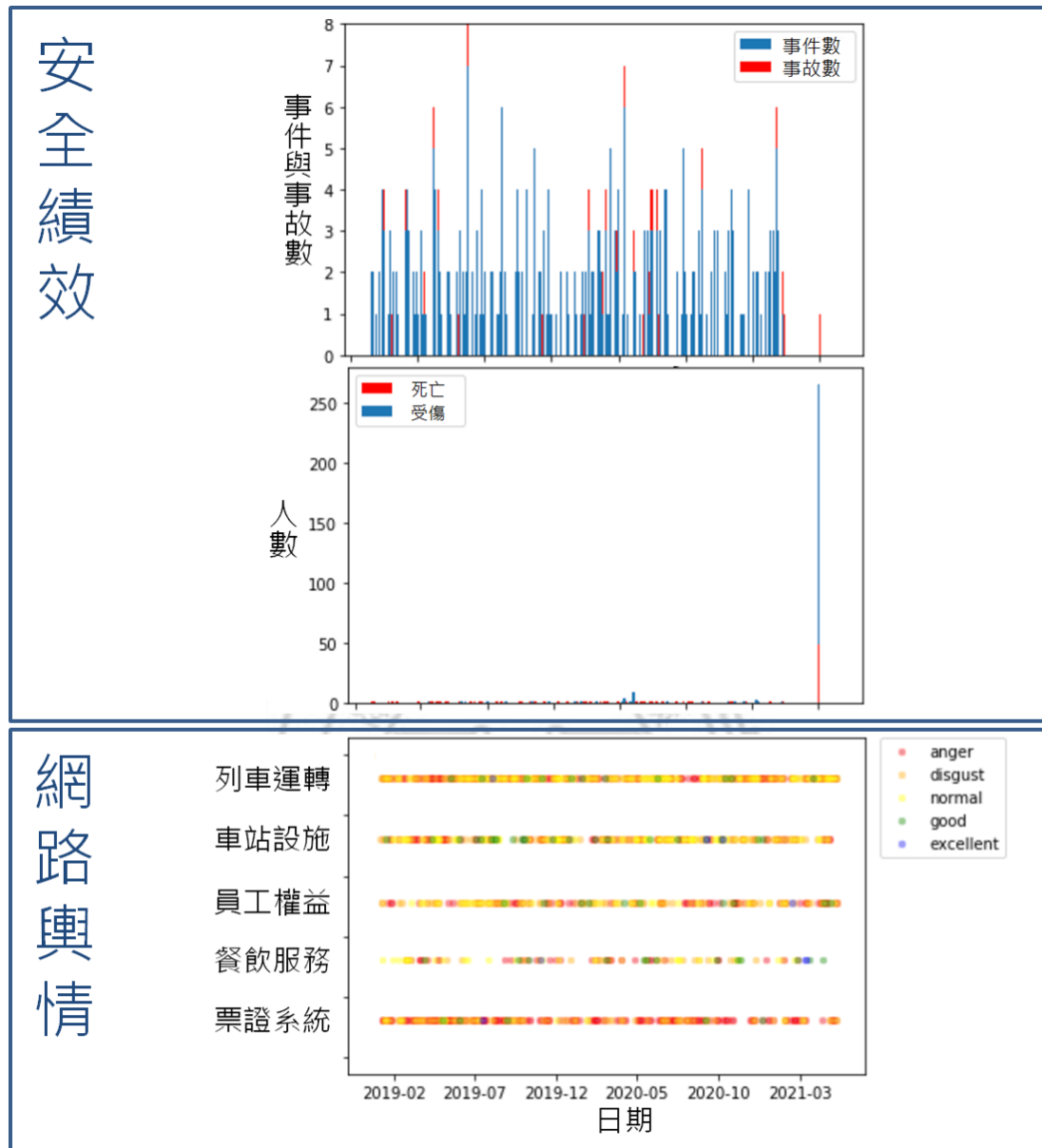


圖 4.25 安全績效與網路輿情對照圖

圖 4.25 為安全績效與網路輿情對照圖(註：臺鐵行車事故報表資料只統計到 2021 年 1 月)。安全績效的部份為臺鐵行車事故報表資料，藍色代表事件數，紅色代表事故數。網路輿情的部份是 PTT 留言的情感趨勢，紅色表示憤怒，藍色表示滿意。以時間軸檢視安全績效與網路輿情資料，探討相同時間點兩者資料的相關聯，也就是當事件或事故發生時，是否有人員死傷，並且會對網路輿情造成何種的情感變化。舉例來說，當 2021 年 4 月 2 日發生太魯閣號出軌事故造成重

大傷亡人數，同時事故發生造成下方五大類網路輿情產生何種的情感變化，最直接影響的就是列車運轉這個類別的留言，事故發生當日列車運轉類別的網路輿情，就由前一天的綠色轉為橘色，至於為什麼沒有達到紅色的憤怒情緒，原因是太魯閣號事故造成重大傷亡，成為臺鐵半世紀來最嚴重的意外事故，許多民眾在下方留言表示哀悼，而這些留言的情感值大都偏向中立，因此沒有讓該日的留言情感達到紅色的非常不滿意。

該圖是以日期作為情感分類的依據基礎，但是當有事件發生時網路輿情的飆漲是非常迅速，而且可能會持續延燒好幾天，此時可參考 4.3 節各項服務類別的留言數與情感分數折線圖，即可得知單一事件的討論熱度，也可了解該事件所造成的情感趨勢變化。

本研究原本想研究服務績效與網路輿情之關聯模式，服務績效有包含安全、效率、環保等等的指標項目，但礙於資料取得困難，最後本研究只以安全績效為例來對照網路輿情。因臺鐵目前只有提供行車事故報表資料，未來若有更多的資料來源，例如：車站營運狀態、票箱收入資料、便當銷售數據、局長信箱旅客意見等資料，則可與不同的服務績效進行對照，進而建立服務績效與使用者體驗結合之整體視覺化儀表板，以使網路輿情關聯模型分析更加全面。

4.6 管理意涵

根據表 4.2 之統整結果，可看出五類評論只有列車運轉類別在關聯模型中最具有顯著性。平均情感分數最低的評論類別為票證系統，餐飲服務雖看來表現最好，但平均情感分數仍為負數，顯示臺鐵各項服務皆有很大改善空間。

本研究旨在提出一套應用於軌道運輸業者服務評論之社群媒體意見挖掘與安全績效對應之流程，首先藉由收集 Ptt 實業坊 Railway 版上有關臺鐵相關貼文之評論建立社群媒體資料庫，然後使用視覺語意深度學習方法進行該社群媒體資料庫評論之意見挖掘，最後借助視覺化技術完成一套安全績效與網路輿情對照顯示介面。

網路評論是旅客對體驗品質的體現，營運管理者可觀察各個服務類別評論情感意向的時空變化趨勢，了解目前各項服務的績效狀態，亦可作為軌道營運業者提升安全及服務品質的一環。營運管理者對於新出現的負面評論可進行即時檢視，建立因應負面評論的回饋機制，將負面評論視為危及品質管理的事件，依據所屬內容分類指派該對應之負責單位立即進行內容檢視，並研擬改善措施。

表4.2 五大類評論結果統整

評論類別	平均情感分數	關聯模型顯著性
車站設施	-0.35361	不具顯著性
員工權益	-0.40947	不具顯著性
餐飲服務	-0.27120	不具顯著性
票證系統	-0.62361	不具顯著性
列車運轉	-0.44260	具有顯著性

第五章 結論與建議

5.1 結論

在考量研究範圍、評論文本資料、情感分析以及關聯模型之情形下，本研究以臺鐵為本體的社群媒體評論資料庫部分，包括資料取得、透過網路爬蟲取得評論文本等工作；在臺鐵服務評論之意見挖掘分類模式部分，包括資料預處理、模型建立、情感值計算等工作；最後在意見挖掘視覺化部分，則包括安全績效與網路輿情關聯模型、成果視覺化。然而此研究成果僅代表部份民眾之真實意見，可供決策者了解問題所在。本研究歸納出以下重要結論及建議：

1. 本研究透過網路爬蟲收集自 2019 年 1 月至 2021 年 5 月期間，於 Ptt 實業坊 Railway 版上有關臺鐵的相關貼文，總共蒐集 2054 篇貼文，84779 則留言，並將評論主題分為「車站設施」、「員工權益」、「餐飲服務」、「票證系統」、「列車運轉」五大類相關文本。
2. 五大類評論主題，平均情感分數最高的為餐飲服務類，最低的為票證系統類，且五個類別的平均分數皆為負數，顯示臺鐵各項服務皆有很大改善空間。
3. 安全績效與網路輿情模型使用 SAS 軟體進行相關係數計算，結果顯示五大類評論中，只有列車運轉類別是具有顯著性的，表示事故或事件的發生會影響列車運轉類別的情感分數，發生事故越多，情感分數就越低。另外，留言情感總分數與總留言數呈高度負相關，相關係數 r 值為-0.98445， P 值小於 0.0001 代表具有顯著性，表示留言數量多寡會影響留言情感總分，且五大類別評論皆是，留言數越多，情感分數就越低，符合網路輿情貶多於褒之特性。
4. 本研究將每日評論情感分類為五個類別，包括：非常不滿意(anger)、不滿意(disgust)、普通滿意(normal)、很滿意(good)、非常滿意(excellent)。其中情感

傾向以非常不滿意型數量最多高達 529 次，非常滿意型最少只有 8 次，反映整體旅客對臺鐵服務的情感傾向非常不滿意。

5. 網路評論是旅客對體驗品質的體現，營運管理者可觀察各個服務類別評論情感意向的時空變化趨勢，了解目前各項服務的績效狀態，亦可作為軌道營運業者提升安全及服務品質的一環。
6. 本研究借助視覺化技術，將安全績效與網路輿情兩者合併成對照圖，以時間軸檢視安全績效與網路輿情資料，探討相同時間點兩者資料的相關聯，亦即當事件或事故發生時，可直觀網路輿情的變化。
7. 臺鐵目前提供旅客意見反應的管道，僅有於車站、網站設置設置站長信箱與客服電話，由電話或網路接受民眾抱怨並作訪談紀錄，再轉由專人處理，並將處理情形向民眾回報。此一作法不僅費工費時，亦無法即時掌握民眾需求，建議未來臺鐵可直接於官方「臺鐵 e 訂通」APP 或官網新增旅客意見專區，讓民眾可用最便民的方式反應意見。
8. 臺鐵目前僅一年一次委外執行整體旅客服務滿意度調查，此一調查方式無法即時反應旅客對於服務體驗的滿意度，亦無法得知各項服務的績效狀態。網路評論亦是旅客對於體驗品質的展現，建議未來臺鐵可透過觀察各項服務評論的網路情感意向趨勢變化，以了解目前各項服務的績效狀態，亦可對於新出現的負面評論進行即時檢視，建立因應負面評論的回饋機制，研擬改善措施，可作為臺鐵提升整體服務品質的輔助工具。

5.2 建議

1. 目前本研究所建置之爬蟲系統抓取對象係以社群媒體 Ptt 實業坊網站為主，考慮未來應用的擴展與分析面向的增加，社群媒體文本的來源亦可持續加廣，如：Mobile 01、Facebook、Instagram、Twitter 等大型討論網站下之留言內容進行抓取，以利分析內容更加全面。同時亦建議應以固定時間(每天、每月、每週)定期爬蟲抓取網路評論，將新評論收錄至資料庫並進行情感分析。
2. 礙於資料取得困難，本研究係以臺鐵行車事故報表資料對應網路輿情資料，並將網路輿情資料分為五大類，然結果不甚理想，未來臺鐵若有新的統計報表資料，可與不同類別的網路輿情資料搭配作分析，如：車站票箱營收、便當銷量統計、旅客投訴項目等，以使網路輿情關聯模型分析內容更加全面。
3. 本研究雖已建立臺鐵輿情情感分析模式，可做為相關領域研究參考，然臺鐵服務範圍廣大，各服務領域下所之輿情評論文本是否具可分析性，仍需進行嘗試，若輿情評論文本數量過少，則較無分析價值。

參考文獻

1. Das, S. and Chen, M. (2001).“Yahoo! for Amazon : Extracting Market Sentiment from Stock Message Boards”, Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), pp.1367-1373.
2. Dave, K., Lawrence, S., and Pennock, D. M.(2003), “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews”, Proceedings of WWW, pp.519-528.
3. Chan, S. M., Cho, M., and Lee, S.(2013), “User Perceptions of Social Media: A Comparative Study of Perceived Characteristics and User Profiles by Social Media”, Journal of Communication and Media Technologies, pp. 149-178.
4. Wilson, K., Fornasier, S., and White, K. M.(2013), “Psychological Predictors of Young Adults' Use of Social Networking Sites”, Cyberpsychology, Behavior and Social Networking, Vol. 13, pp. 173-177.
5. Eva Nedeliaková, Jana Sekulová, Ivan Nedeliak, Martin Loch.(2014). Methodics of identification level of service quality in railway ransport. Procedia - Social and Behavioral Sciences,110,320-329.
6. José Luis Machado-Leóna , Rocío de Oña, Tahar Baounib ,Juan de Oña.(2017). Railway transit services in Algiers: priority improvement actions based on users perceptions. Transport Policy,53,175-185.
7. Yan, Z., Jing, X., & Pedrycz, W. (2017). Fusing and mining opinions for reputation generation. Information Fusion, 36, 172-184.
8. Wencheng Huanga , Bin Shuai , Yan Sun , Yang Wang , Eric Antwi.(2018). Using entropy-TOPSIS method to evaluate urban rail transit system operation performance: The China case. Transportation Research Part A: Policy and

- Practice,111,292-303.
9. Francisco Díez-Mesa, Rocio de Oña, Juan de Oña.(2018).Bayesian networks and structural equation modelling to develop service quality models: Metro of Seville case study. *Transportation Research Part A: Policy and Practice*,118,1-13.
 10. Heike Link.(2019).The impact of including service quality into efficiency analysis: The case of franchising regional rail passenger serves in Germany. *Transportation Research Part A: Policy and Practice*,119,284-300.
 11. Luis Oliveira, Claudia Bruen, Stewart Birrell, Rebecca Cain.(2019). What passengers really want: Assessing the value of rail innovation to improve experiences. *Transportation Research Interdisciplinary Perspectives*,1,100014.
 12. Madan Lal Yadav, Basav Roychoudhury.(2019).Effect of trip mode on opinion about hotel aspects: A social media analysis approach. *International Journal of Hospitality Management*,80,155-165.
 13. Jose Agustin Vallejo-Borda, Daniel Rosas-Satizabal,Alvaro Rodriguez-Valencia.(2020).Do attitudes and perceptions help to explain cycling infrastructure quality of service? *Transportation Research Part D: Policy and Practice*,87, 102539.
 14. Ivo Sousa, Maria Paula Queluz, Antonio Rodrigues.(2020). A survey on QoE-oriented wireless resources scheduling. *Journal of Network and Computer Applications*,158, 102594.
 15. Gabriela Gongora Svartzman, Jose E. Ramirez-Marquez, Kash Barker.(2020). Social media analytics to connect system performability and quality of experience, with an application to Citibike. *Computers & Industrial Engineering*,139, 106146.

16. Jyoti Mandhani , Jogendra Kumar Nayak, Manoranjan Parida.(2020).
Interrelationships among service quality factors of Metro Rail Transit System:
An integrated Bayesian networks and PLS-SEM approach. Transportation
Research Part A: Policy and Practice,140,320-336.
17. Shan-sheng ZHU, Yu-ning DONG, Cheng XU.(2020).A Statistical QoE-QoS
Model of Video Streaming Services. 2020 The 6th International Conference on
Computing and Data Engineering,195-199.
18. Liming Wu, Haishan Xia, Xianfeng Wang, Qingfeng Dong, Chunxiang Lin,
19. Xiaotong Liu, Rui Liang.(2020).Indoor air quality and passenger thermal
comfort in Beijing metro transfer stations. Transportation Research Part D,78,
102217.
20. Joaquín Osorio-Arjona, Jiri Horak, Radek Svoboda, Yolanda García-Ruíz.(2021).
Social media semantic perceptions on Madrid Metro system: Using Twitter data
to link complaints to space. Sustainable Cities and Society,64,102530.
21. Fajar Fathur Rachmana , Rani Nooraenia , Lia Yuliana.(2021). Public Opinion of
Transportation integrated (Jak Lingko), in DKI Jakarta, Indonesia. Procedia
Computer Science 179, 696–703.
22. 李治綱、鍾志成、林杜震、張仕龍、張恩輔、陳一昌、張開國、吳熙仁
(2009)，公共運輸之安全績效：臺灣鐵路管理局之個案分析。運輸計劃季
刊，38(4)期，381-405。
23. 蕭穎謙(2009)，服務品質、顧客滿意度與忠誠度關係之研究—以臺灣高速
鐵路股份有限公司為例。全球管理與經濟期刊，5(2)，50-61。
24. 賴淑芳(2012)，公共自行車接受度與滿意度研究—以臺北微笑單車為例。運
輸學刊，24(3)，379-406。

25. 吳吉祥、李凡、夏靖波、王愷(2012)，基於廣義線性模型的QoE評估方法研究。計算機應用研究，29(10)。
26. 周瑞生、葉竣賓(2013)，從旅客觀點探討影響高速鐵路營運績效購念暨其中介與調節效果。營建管理季刊，94，1-23。
27. 劉歡、高月紅、林言超、桑林、楊大成(2014)，一種研究QoE與QoS關係的數學模型方法。廣東通信技術，34(10)。
28. 施伯燁(2014)，社群媒體—使用者研究之概念、方法與方法論初探。傳播研究與實踐，4(2)，207-227。
29. 朱斌妤、黃東益、洪永泰、李仲彬、曾憲立(2015)，「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。
30. 游綉雯(2015)，使用情緒分析於社群論壇消費者評論滿意度評估之研究—以TripAdvisor旅遊網站為例。中興大學行銷學系所碩士學位論文。
31. 陳亭愷(2015)，社群運算應用於網路輿情情感傾向分析之研究—以實施國道計程電子收費政策為例，淡江大學運輸管理學系碩士論文。
32. 陶冶中、簡睿志(2016)，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。運輸計劃季刊，45(4)，301-330。
33. 蔡易辰(2016)，三元決策理論應用於國道計程收費議題之情感分析研究，淡江大學運輸管理學系碩士論文。
34. 陳克罕(2016)，臺灣高鐵行車事件風險管理之研究，逢甲大學運輸科技與管理學系碩士在職專班碩士論文。
35. 朱連章、李然然、張紅霞、郭加樹、張泉(2017)，基於QoS 量化用戶體驗質量的評價模型。計算機系統應用，26(5)。
36. 洪敬政(2017)，情感分析研究的知識結構，淡江大學資訊管理學系碩士在職專班學位論文。

37. 趙玉娟與陶振超(2018)，政治網路口碑的情感分析：語意關連性之觀點。
傳播研究與實踐，8(2)，75-107。
38. 陳翰(2018)，從社群媒體挖掘以感測日常交通滿意度之研究，淡江大學運輸管理學系碩士論文。
39. 林彥伯(2018)，捷運系統服務指標—以臺中捷運綠線為例，逢甲大學運輸與物流學系碩士論文。
40. 鄭永祥(2019)，臺鐵普悠瑪列車事故之反思。消費者報導雜誌，453期，P25-30。
41. 楊竑昕(2019)，運用自然語言處理工具實作貼文分析系統觀察網路論壇，以 Dcard為例，國立中山大學資訊管理研究所碩士論文。
42. 張悅朗(2020)，視覺語意深度學習方法應用於社群媒體意見挖掘之研究-以軌道運輸場站服務評論為例，淡江大學運輸管理學系碩士論文。
43. 簡信立(2020)，鐵路運輸緊急應變處理之研析-以臺鐵局人為疏失出軌事故為例，成功大學交通管理科學系碩士在職專班學位論文。
44. 徐倩文、韋健華、吳忠宏、董二為(2020)，休閒產業管理學刊，13(1)，51 - 62。
45. 徐任宏、林杜寰、孫千山、張開國、葉祖宏、吳熙仁(2021)，國內外鐵道安全保證機制之回顧與展望。中興工程，150期，21-30。
46. 國家運輸安全調查委員會(2020)，1021臺鐵第6432次車新馬站重大鐵道事故-補強-調查報告-第一冊，國家運輸安全調查委員會(編號：TTSB-ROR-20-10-001)。
47. 國家運輸安全調查委員會(2020)，1021臺鐵第6432次車新馬站重大鐵道事故-補強-調查報告-第二冊，國家運輸安全調查委員會(編號：TTSB-ROR-20-10-001)。

48. 林杜寰、孫千山、李治綱、陳桂豪、吳明軒、胡仲瑋、張開國、蔡祖宏、吳熙仁、洪憲忠(2019)，鐵路運輸安全管理系統(SMS)制度化策略之研擬，交通部運輸研究所。
49. 孫千山、鍾志成、李治綱、薛強、林杜寰、張仕龍、張恩輔、林蓁、黃笙玹、李永強、陳一昌、張開國、賴靜慧、吳熙仁(2011)，風險管理應用於鐵路運輸安全之初探-以臺鐵風險辨識為例，交通部運輸研究所。
50. 孫千山、鍾志成、李治綱、林杜寰、張仕龍、張恩輔、林蓁、黃笙玹、黃宏仁、張開國、賴靜慧、吳熙仁(2012)，風險管理應用於鐵路運輸安全之初探-以臺鐵風險分析與評量為例，交通部運輸研究所。
51. 孫千山、鍾志成、李治綱、林杜寰、張仕龍、張恩輔、林蓁、黃笙玹、黃宏仁、張開國、賴靜慧、吳熙仁(2013)，風險管理應用於鐵路運輸安全之初探-以臺鐵風險處理、管理監督、管理改善為例，交通部運輸研究所。
52. 孫千山、鍾志成、李治綱、林蓁、施佑林、吳明軒、張開國、葉祖宏、賴靜慧、吳熙仁(2016)，鐵路安全之風險管理推動研究—發展鐵路系統之安全管理實務與報告，交通部運輸研究所。
53. 國家運輸安全調查委員會(2019)，重大鐵道事故通報表，擷取自2021/05。
<https://www.tsb.gov.tw/media/3280/marine-occurrence-report-form.pdf?mediaDL=true>
54. 國家運輸安全調查委員會(2019)，運輸事故調查法，擷取自2021/05。
<https://www.tsb.gov.tw/1133/1206/25253/25260/25296/post>
55. 國家運輸安全調查委員會(2020)，重大鐵道事故調查作業處理規則，擷取自2021/05。
<https://www.tsb.gov.tw/1133/1206/25253/25260/25305/post>
56. 中華民國交通部網站，鐵路行車規則，擷取自2021/05。
<https://motclaw.motc.gov.tw/webMotcLaw2018/Law/ArticleContent?type=->

1&LawID=D0006010

57. 臺灣鐵路管理局網站，事故資料月報表，擷取自2021/05。
<https://www.railway.gov.tw/tra-tip-web/adr/about-public-info-6-2>
58. Ptt實業坊Railway版，擷取自2021/05。
<https://www.ptt.cc/bbs/Railway/index.html>
59. 中央研究院(2020)， CkipTagger 中文說明，擷取自2021/05。
<https://github.com/ckiplab/ckiptagger/wiki/Chinese-README>
60. PPT情感分析，擷取自2021/05。GitHub - kuo23/PTT-Stock
61. GitHub - PPT股票版爬文與情感分析，擷取自2021/05。
<https://github.com/wellgary/PTT-Crawler-and-Emotional-Analysis/blob/master/PTT%20Crawler%20and%20Emotion%20Analysis.ipynb>
62. 讓電腦聽懂人話: 直觀理解 Word2Vec 模型，擷取自2021/05。
<https://tengyuanchang.medium.com/%E8%AE%93%E9%9B%BB%E8%85%A6%E8%81%BD%E6%87%82%E4%BA%BA%E8%A9%B1%E7%90%86%E8%A7%A3-nlp-%E9%87%8D%E8%A6%81%E6%8A%80%E8%A1%93-word2vec-%E7%9A%84-skip-gram-%E6%A8%A1%E5%9E%8B-73d0239ad698>
63. 相關係數與共變異數(Correlation Coefficient and Covariance) ，擷取自2021/05。
<https://chih-sheng-huang821.medium.com/%E7%9B%B8%E9%97%9C%E4%BF%82%E6%95%B8%E8%88%87%E5%85%B1%E8%AE%8A%E7%95%B0%E6%95%B8-correlation-coefficient-and-covariance-c9324c5cf679>
64. 統計學: 皮爾森相關係數為什麼小於等於1，擷取自2021/05。
<https://chih-sheng-huang821.medium.com/%E7%B5%B1%E8%A8%88%E5%AD%B8->

%E7%9A%AE%E7%88%BE%E6%A3%AE%E7%9B%B8%E9%97%9C%E4
%BF%82%E6%95%B8%E7%82%BA%E4%BB%80%E9%BA%BC%E5%B0
%8F%E6%96%BC%E7%AD%89%E6%96%BC1-180de90df79d

65. 皮爾森積動差相關係數，維基百科，擷取自2021/05。

[https://zh.wikipedia.org/zh-](https://zh.wikipedia.org/zh-tw/%E7%9A%AE%E5%B0%94%E9%80%8A%E7%A7%AF%E7%9F%A9%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0)

[tw/%E7%9A%AE%E5%B0%94%E9%80%8A%E7%A7%AF%E7%9F%A9%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0](https://zh.wikipedia.org/zh-tw/%E7%9A%AE%E5%B0%94%E9%80%8A%E7%A7%AF%E7%9F%A9%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0)

66. DNN — 深度神經網路，擷取自2021/05。

[https://medium.com/%E4%B8%80%E4%BA%BA%E5%A4%9A%E5%B7%A5](https://medium.com/%E4%B8%80%E4%BA%BA%E5%A4%9A%E5%B7%A5%E5%B7%A5%E4%BD%9C%E5%AE%A4/dnn-%E6%B7%B1%E5%BA%A6%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-cf892cbb06d5)

[%E5%B7%A5%E4%BD%9C%E5%AE%A4/dnn-](https://medium.com/%E4%B8%80%E4%BA%BA%E5%A4%9A%E5%B7%A5%E5%B7%A5%E4%BD%9C%E5%AE%A4/dnn-%E6%B7%B1%E5%BA%A6%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-cf892cbb06d5)
[%E6%B7%B1%E5%BA%A6%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-cf892cbb06d5](https://medium.com/%E4%B8%80%E4%BA%BA%E5%A4%9A%E5%B7%A5%E5%B7%A5%E4%BD%9C%E5%AE%A4/dnn-%E6%B7%B1%E5%BA%A6%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-cf892cbb06d5)

67. 深度神經網路(DNN)模型與前向傳播演算法，擷取自2021/05。

<https://www.cnblogs.com/pinard/p/6418668.html>

68. 利用SnowNLP 訓練自己的情感分析庫，擷取自2021/05。

<https://www.itread01.com/content/1541222413.html>

69. [Python] [snownlp]基於情感詞典的情感分析，擷取自2021/05。

<http://corina.cc/article/217/>