

中華民國第 60 屆中小學科學展覽會 作品說明書

高級中等學校組 電腦與資訊學科

佳作

052507

應用網路爬蟲於社交軟體實現群眾互動平臺之
研究

學校名稱：臺北市立松山高級工農職業學校

作者： 職三 周駿曄 職三 楊立鼎 職三 李品宏	指導老師： 余耀銘 林敬堯
---	-----------------------------

關鍵詞：社交軟體、網路爬蟲、圖像辨識

摘要

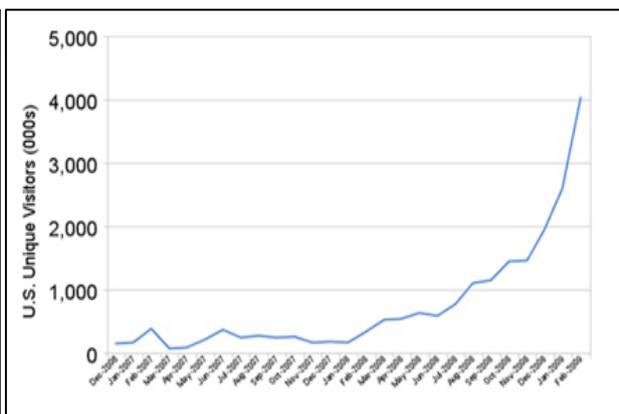
現今的大型活動，如校慶活動、新北耶誕城等，缺乏互動性與參與感，其中原因大多是群眾埋沒於手機的社交軟體所導致。而本研究將利用其便利與群眾互動特性，**使用問卷抽查法探討大眾對於活動的觀點，應用網路爬蟲技術抓取貼文**，設計一套能提高互動性的平臺，使用者只需在 Instagram、Twitter 等社交軟體中發布文章，系統會即時推播至活動中的大螢幕上，並且使用 LineBot 結合圖像辨識自動且快速審核貼文，另一方面以後臺系統，方便活動單位常駐貼文輪播活動訊息、廣告與管理。綜上所述，本研究結果確實有效地提升活動互動性及樂趣，以時下流行的各種社交軟體，配合研究平台為各類大型活動量身打造，也能夠運用於政令宣導或文宣廣告等用途上，大幅提升活動的互動性與參與感。

壹、研究動機

現今因應智慧型手機的普及化，加上 LINE、IG 等軟體能夠即時傳送訊息，因為可以在不額外增加費用之情形下與其他使用者即時傳送文字、圖片等多媒體訊息，受到許多使用者的青睞，自 2016 年 3 月以來，LINE 於全球就擁有超過 10 億人使用，其中更有高達 2.15 億活躍使用者；而 Facebook、Instagram 等等社交應用軟體上，用戶可以在各大平台上分享生活上的相片及視訊，也能關注其他用戶的貼文，這樣的功能受到眾多使用者的喜愛，如【圖一】、【圖二】所示，但這也造成活動當中群眾大多時間都在低頭滑手機以及發文來分享現在的情緒以及感想，使台下與台上間的互動愈來愈冷漠，讓活動中的趣味性及互動性降低。因此我們希望可以研發出一套系統，能將 LINE 快速傳播訊息的特性，以及結合各大社交平台即時分享多媒體訊息的優點，設計出一套系統能將群眾所發布的貼文即時轉貼至舞台上的大螢幕系統當中，希望能夠將群眾的目光從手機上移至活動上，藉此改善群眾在活動中的互動性，讓活動更加的熱絡。使用者在社交平台上發佈的貼文能同時發佈至系統中，也可經由掃描活動單位提供的 QRcode 後進行訊息的發送，活動管理單位也能利用建立於網頁上的線上管理系統來進行訊息內容的控管，使眾多的訊息能快速且正確的顯示於開放空間的螢幕中，進而使大眾可輕易的閱讀到發信者欲傳達的訊息，且我們希望讓此研究不僅能應用於大型活動中，更能使用於各類文宣廣告、政令宣導上。



【圖一】Instagram 活躍用戶統計圖



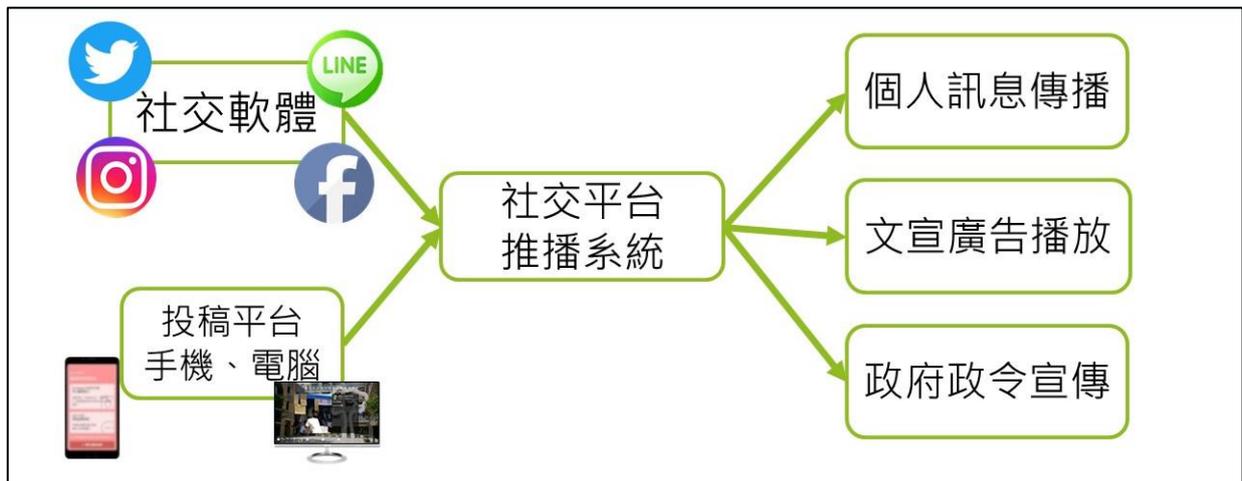
【圖二】Twitter 全球用戶統計圖

貳、研究目的

本作品主要的構想是想創造一個開放的訊息傳遞空間，讓每個人不僅能即時傳遞訊息，也能在推播螢幕看到所有的訊息。在設計的規劃上，使用者可以經由在社交平台貼文並在文章裡加入特定文字，上傳完畢便於顯示器顯示貼文，如此就可以將訊息及時轉貼至活動中的大螢幕上，以增加群眾對活動的參與感。本作品利用現今流行的社交軟體，使用者在發佈貼文的同時能讓系統擷取貼文，通過審核後立即顯示於螢幕上，如此可以方便使用者不需二度發文，有效減少使用者與管理者的不便，讓使用者更有意願投稿以融入於活動當中，如【圖三】所示。

以下是本研究的主要研究目的：

- 一、研究如何提升活動的互動性
- 二、探討不同的社交軟體使用人數比例
- 三、探討如何加快審核速度
- 四、探討網路爬蟲策略比較
- 五、探討使用互動平台前後的互動性比較



【圖三】系統目標

參、研究設備及器材

一、硬體、軟體及作業系統統整表

如【表一】所示為本研究所使用到的各項設備及作業系統表。

【表一】研究設備統整表

硬體設備	開發軟體	伺服器系統
伺服器主機	PHP	Ubuntu
智慧型手機	Line Bot	MySQL
大型螢幕	Microsoft Azure	SSH Server

二、硬體設備介紹

(一) 伺服器主機

為使系統得以快速執行任務，我們採用如【表二】所示的主機規格，使系統在活動當中可提供即時抓取、審核並同步貼文至大型螢幕上。

【表二】伺服器主機規格表

名稱	型號
處理器 CPU	Intel i7-9700K 3.6GHz
主機 MB	華碩 TUF Z390-PLUS GAMING WIFI
記憶體 RAM	金士頓 16GB*2

固態硬碟 M.2SSD	三星 Samsung 860 PRO 256G
傳統內接硬碟 HDD	WD 1TB
CASE 機殼(+電源)	酷碼 MasterBox MB511
電源供應器	台達 550W

(二) 智慧型手機

為測試使用者在活動當中發布貼文時系統是否能正確抓取，我們以智慧型手機發布貼文做為測試，如【圖四】所示為智慧型手機示意圖。



【圖四】智慧型手機示意圖

(三) 大型螢幕

大型螢幕可顯示在活動當中的觀眾貼文，讓大眾的心情得以被其他人看見，以此來增加活動的互動性與參與感，如【圖五】所示為大型螢幕示意圖。



【圖五】大型螢幕示意圖

三、開發軟體介紹

(一) PHP

PHP 為 PHP: Hypertext Preprocessor 的縮寫，PHP 是一種跨平台、嵌入式的伺服器端執行的 script language。適合於網頁的開發，並可內嵌入 HTML 中。PHP 主要功能在於伺服器端的 script 程式，因此可以用它來進行任何 CGI 程式可以做到的事，例如收集表單資料、動態產生網頁或者發送/接收 Cookies，如【圖六】所示。



【圖六】PHP

(二) Line Bot

Line BotAPI 提供與顧客間的雙向互動，透過 LINE 介面就能享受店家各項服務及活動，還能串接資料庫，讓所有服務及資訊個人化，如同每位顧客專屬的線上專員，如【圖七】所示。



【圖七】LINE BOT

(三) Microsoft Azure

Azure 是開放且彈性的雲端平台，可讓開發者跨 Microsoft 管理的全球性資料中心網路快速建置、部署及管理應用程式。開發者可以使用任何語言、工具或架構來建立應用程式，並可將公用雲端應用程式與現有的環境進行整合，如【圖八】所示。



【圖八】Microsoft Azure

四、伺服器系統介紹

(一) Ubuntu

Ubuntu 是基於 Debian 發行版和 GNOME 桌面環境。Ubuntu 的目標在於為一般用戶提供一個最新同時又相當穩定，主要以自由軟體建構而成的作業系統。Ubuntu 目前具有龐大的社群力量支援，用戶可以方便地從社群獲得幫助，如【圖九】所示。



【圖九】Ubuntu

(二) MySQL

MySQL 是一種開放源代碼的關係型數據庫管理系統 (RDBMS)，MySQL 數據庫系統使用最常用的數據庫管理語言：結構化查詢語言 (SQL) 進行數據庫管理。由於 MySQL 是開放源代碼的，因此任何人都可以在通用公共許可證的許可下載，並根據個性化的需要對其進行修改，如【圖十】所示。



【圖十】MySQL

(三) SSH Server

SSH Server 可讓管理員不管發生任何狀況，或是要進行系統更新，都不需要回到主機前面操作系統即可遠端完成。若是發生資安事件也可立即做出反應，將損失降至最低，如【圖十一】所示。



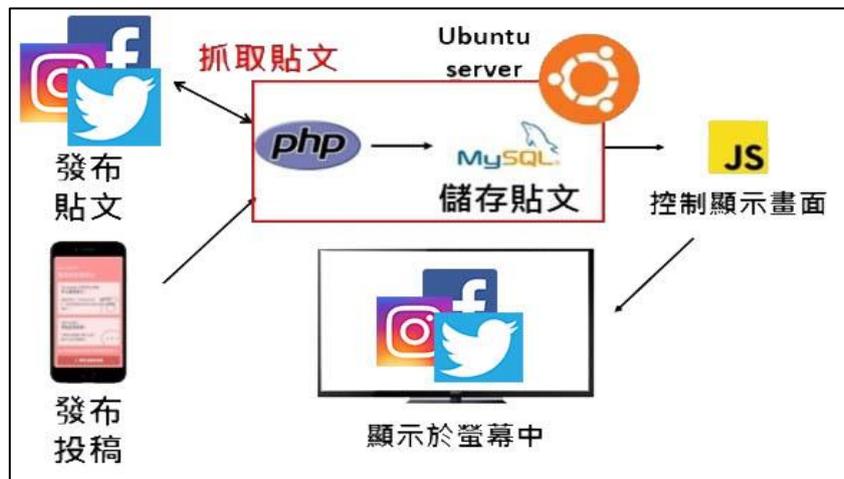
【圖十一】SSH Server

肆、研究過程或方法

一、系統規劃

(一) 系統架構

本作品發想於現今的廣告聯播網並進行改良。使用者可於社交平臺發文或至網頁投稿，發布的貼文會經由我們所設計的網路爬蟲進行抓取，並儲存至伺服器中的 MySQL 資料庫。抓取後的貼文會經由 LINE BOT 自動審核，當貼文成功通過審核的同時，將同步至推播平台。作品相較於現今的聯播網系統更有即時性與便利性，如【圖十二】所示為系統運作流程圖。



【圖十二】系統架構圖

(二) 系統特色

本作品希望未來不僅能應用於大型活動中，更能應用在大型商業廣告、政令宣導等不同用途上，而依此發想的研究四大特色，如【圖十三】所示。目前我們讓使用者能在 Instagram、Twitter 發布貼文同時讓貼文顯示於系統中，原先是要做 Facebook 爬蟲但因先前發生了嚴重的個資外洩問題，導致 Facebook 已經將大量的 API 關閉，所以無法利用 API 抓取貼文。雖然如此，我們還是讓不熟悉上述兩個社交平臺的使用者也能透過特定網站進行投稿，即時將多媒體訊息顯示於大螢幕上，進而達到群眾互動之效果。而為了避免不當言論及不雅圖片顯示於公眾場合，我們利用 LINE 其中的群組功能配合 LINE BOT 功能，可以讓多位管理員在群組中同時進行審核，如此能加速審核訊息更能避免有大量貼文湧入時卻只有一位管理員審核訊息的窘境，

同時我們也加入**圖像辨識自動審核**使審核速度更加快速。訊息通過審核之後便會加入下一輪的輪播中，以**輪播**的方式能快速瀏覽不同的貼文，也可以設定特定貼文不會被後續上傳的貼文蓋過，達到加強文宣廣告、政令宣導之效果，更能為大型活動量身打造，增加活動的互動性及可看性。



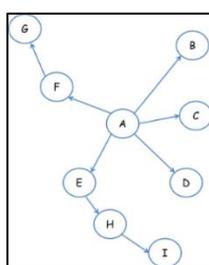
【圖十三】研究特色圖

二、探討網路爬蟲技術

網路爬蟲是一種「自動化瀏覽網路」的程式，或者說是一種網路機器人。他們被廣泛用於網際網路搜尋引擎或其他類似網站，以取得或更新這些網站的內容和檢索方式。他們可以自動採集所有其能夠存取到的頁面內容，以供搜尋引擎做進一步處理，處理成自己想要的格式。

(一) 深度優先策略

深度優先是指網路爬蟲會從起始頁開始，一個分支一個分支挖掘資料，處理完這條分支之後再轉回節點，繼續挖掘下一個分支的資料，並進行深度的搜索，直到該節點所有的分支都已探索，重複探尋未搜索到的資料，直到找到目標資料或是尋遍所有節點與分支，如【圖十四】所示為深度優先策略概念圖。

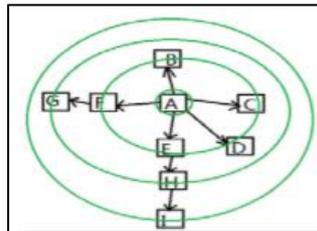


【圖十四】深度優先策略概念圖

(圖十二資料來源：張華明(2018)。基於網路爬蟲和聲控技術應用於線上音樂點歌之研究。聖約翰科技大學：碩士論文。)

(二) 寬度優先策略

寬度優先策略會從某一網站開始走訪，接著走訪此網站所有相鄰且未拜訪過的網站，再從這些網站進行先廣後深的搜尋，當每個網站同一深度的節點走訪完，再繼續向下一個深度搜尋，直到找到目的節點或搜尋全部節點，如【圖十五】所示為深度優先策略概念圖，先搜尋完同個圈內(深度)的資料後，再擴大深度搜索。

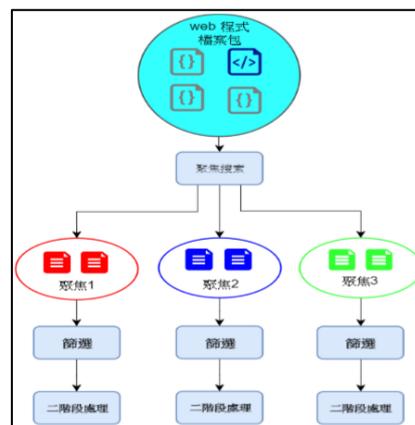


【圖十五】寬度優先策略概念圖

(圖十三資料來源：張華明(2018)。基於網路爬蟲和聲控技術應用於線上音樂點歌之研究。聖約翰科技大學：碩士論文。)

(三) 聚焦搜索策略

聚焦搜索策略是根據「匹配優先原則」對數據進行訪問。基於特定的匹配算法，主動選擇與主題相關的資料，並限定優先級，以引導後續的資料抓取。根據「最好優先原則」進行訪問，讓程式可以快速有效地獲得更多與主題相關的頁面，如【圖十六】所示為深度優先策略概念圖，圖中的「聚焦」為指定資料抓取的主題，抓取後會經過篩選與處理，呈現出使用者所需樣貌。



【圖十六】聚焦搜索策略概念圖

(圖十四資料來源：張華明(2018)。基於網路爬蟲和聲控技術應用於線上音樂點歌之研究。聖約翰科技大學：碩士論文。)

(四) 深度優先、寬度優先、聚焦搜索策略比較

我們的整套系統是利用網路爬蟲抓取使用者的貼文同步顯示於活動中的大螢幕上，增加活動中的互動性以及趣味性。為了有效的抓取使用者的貼文，我們使用了以主題優先搜索且更新效率快的聚焦搜索策略。拿我們所使用的聚焦搜索策略來跟其他兩個搜索策略比較後，發現聚焦搜索策略確實優於其他兩個搜索策略，比如抓取速度快、主題式搜索等。如【表三】所示為三種網路爬蟲策略比較表。

【表三】深度優先、寬度優先、聚焦搜索網路爬蟲策略比較表

比較 \ 搜索策略	深度優先策略	寬度優先策略	聚焦搜索策略
抓取速度	慢	普通	最快
效率	普通	普通	較好
以主題進行搜索	否	否	是



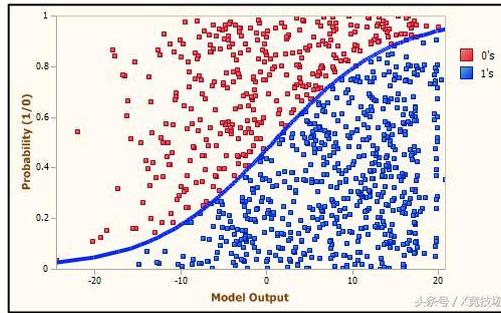
(表三資料來源：自行設計)

三、探討圖像辨識技術

圖像辨識技術可能是以圖像的主要特徵為基礎的，每個圖像都有它的特徵。在人類圖像辨識系統中，對複雜圖像的識別往往要通過不同層次的信息加工才能實現。圖像辨識技術是立體視覺、運動分析、數據融合等實用技術的基礎，在研究等許多領域可廣泛應用。我們希望透過這項技術來判斷貼文中是否有不雅圖片，對貼文進行審核。

(一) 邏輯回歸(Logistic Regression)

相似於線性回歸，同樣是找出自變數(Independent variable)依變數(Independent variable)的關係，然而不同於線性迴歸，邏輯回歸的輸出(output)二進位的，不是 1 即是 0。邏輯回歸為最頻繁被使用的分類演算法，時常被應用於風險評估或判斷異常，如下【圖十七】所示。



【圖十七】邏輯回歸概念圖

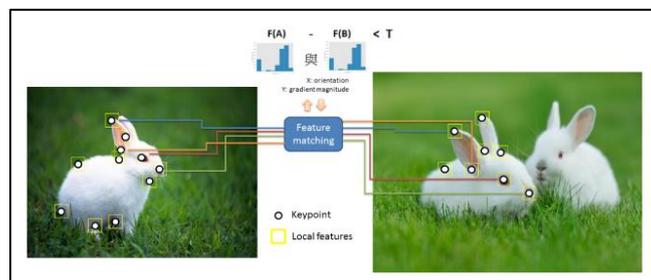
(圖十七資料來源：機器學習 - 定義、演算法、以及商業應用。2020年6月13日，取自

<https://oosga.com/machine-learning/>)

(二) 自適應增強(AdaBoost)

自適應增強整合多種不同的模型來做分類或回歸分析，再根據其正確率去判斷每一個不同的輸出所佔的比重為何。自適應增強以及其他增強(Boosting)模型，在提高準確率上非常有效。應用於低成本的方式來做圖片識別。基於自適應增強的原理，就能設計出特徵點配對(Feature matching)，如【圖十八】所示為利用特徵點配對判斷特徵示意圖，而特徵點配對的判斷步驟如下：

1. 關鍵點偵測(Keypoint detection)：在圖片中取得關鍵點（可能為 edges、corners 或 blobs）。
2. 特徵點提取(Feature extraction)：針對各關鍵點提取該區域的 features(稱為 local features)。
3. 關鍵點篩選並進行特徵點配對。



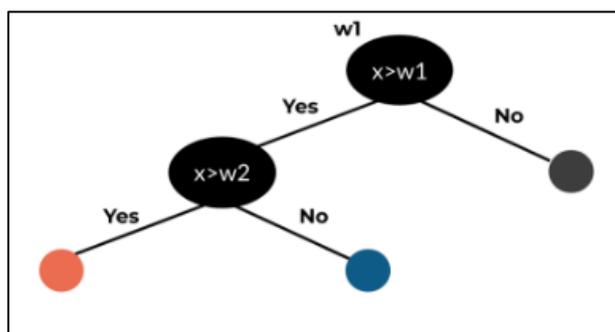
【圖十八】利用特徵點配對(Feature matching)判斷兔子特徵

(圖十八資料來源：圖像特徵比對(一)-取得影像的特徵點。2020年6月13日，取自

<https://reurl.cc/j72gEp>)

(三) 決策樹

決策樹利用不同數據特性的值分成不同的分支，像一棵樹上的樹枝一般。如同銀行在決定是否要貸款給某客戶時，行員會透過一系列的問題以判斷最終決策。決策樹利用同樣的架構，不斷地拆分數據的特性，直到持續分出分枝將不會提供任何價值為止。此種特性也讓決策數在面對高維數據時，依然有可觀的正確率，如【圖十九】所示。



【圖十九】決策樹概念圖

(圖十九資料來源：機器學習 – 定義、演算法、以及商業應用。2020年6月13，取自

<https://oosga.com/machine-learning/>)

(四) 邏輯回歸、自適應增強、決策樹比較

我們研究中的圖像辨識是用來判斷審核貼文時是否有不雅貼文出現，以防止在活動中出現不雅畫面時導致活動發生混亂。經由如【表四】所比較的項目，我們在研究中使用了能利用特徵去判斷圖片的自適應增強。

【表四】邏輯回歸、自適應增強、決策樹辨識方法比較表

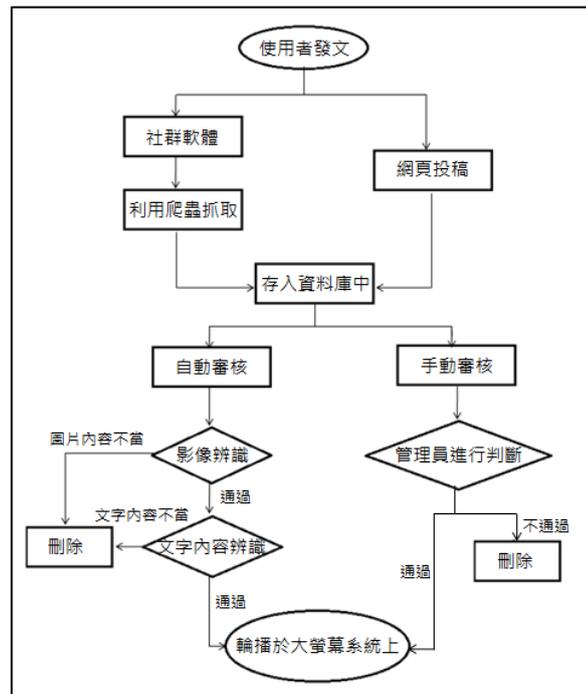
辨識方法 比較	邏輯回歸	決策樹	自適應增強
判斷準確率	中	中	較高
利用圖片的 特徵判斷	否	否	是



(表四資料來源：自行設計)

四、系統流程圖

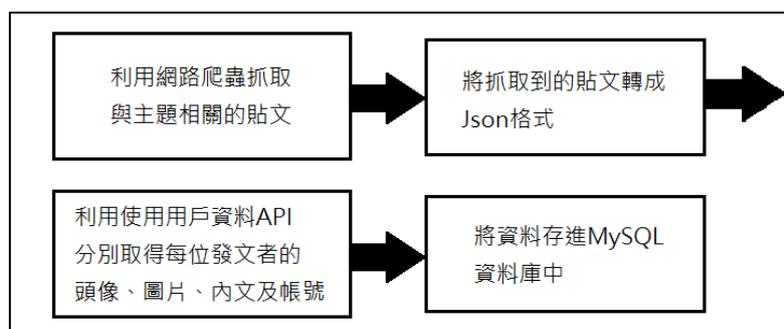
在本研究中我們使用 php 以及 JavaScript 建構整套系統，使用 php 撰寫伺服器端以及網路爬蟲的程式碼，像是網路爬蟲、網路後臺等等功能，利用 JavaScript 顯示輪播的內容。本研究整體運作流程如【圖二十】所示。



【圖二十】系統運作流程圖

(一) 抓取 Instagram 與 Twitter 貼文

我們以 PHP 的 Curl 取得 Instagram 或 Twitter 回傳的 Json 格式資料轉為陣列，並使用用戶資料 API 分別取得每位發文者的頭像以便顯示，最後將取得到的圖片、內文、帳號等資料存進 MySQL 中，如【圖二十一】所示。



【圖二十一】抓取貼文流程圖

(二) 圖像辨識訓練

如【圖二十二】所示為本作品所使用的圖像辨識訓練流程圖，其步驟如下所述：

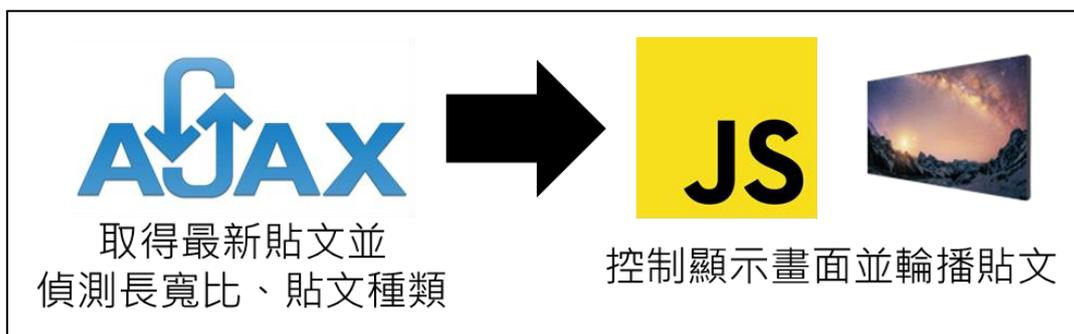
1. 選出多張圖片輸入訓練模型中。
2. 將圖片中人物的特徵描繪出來。
3. 讓機器判斷這些圖片那些能通過哪些不能通過。
4. 再將正確解答輸入到訓練模型中，讓機器更清楚了解如何判斷。
5. 將機器訓練完後，再次輸入多張圖片進行判斷，驗證機器是否能夠正確的判斷出圖片是否有不雅內容。
6. 最後再將訓練模型的判斷過程微調讓機器能夠更準確的審核貼文。



【圖二十二】 圖像辨識訓練流程圖

(三) 電視牆取得資料

這部分我們使用 Javascript 的 interval 在每輪播放結束前以 Ajax 技術取得最新的貼文，並偵測貼文種類與圖片長寬比，自動決定每篇貼文的顯示樣式，如【圖二十三】所示。

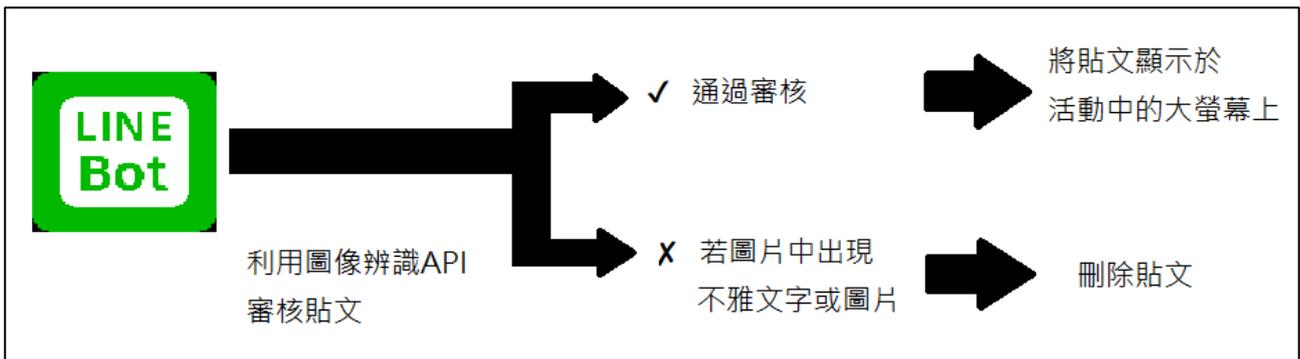


【圖二十三】 電視牆取得顯示資料流程圖

(四) Line Bot 自動過濾不當貼文

我們利用圖像辨識去辨識貼文中是否有不當的文字或圖片，如果發現不當貼文便立即刪除，如果審核通過就將貼文顯示在活動中的大螢幕上，並且搭配 Line Bot

讓活動單位可以輕鬆觀察審核的結果，如【圖二十四】所示。

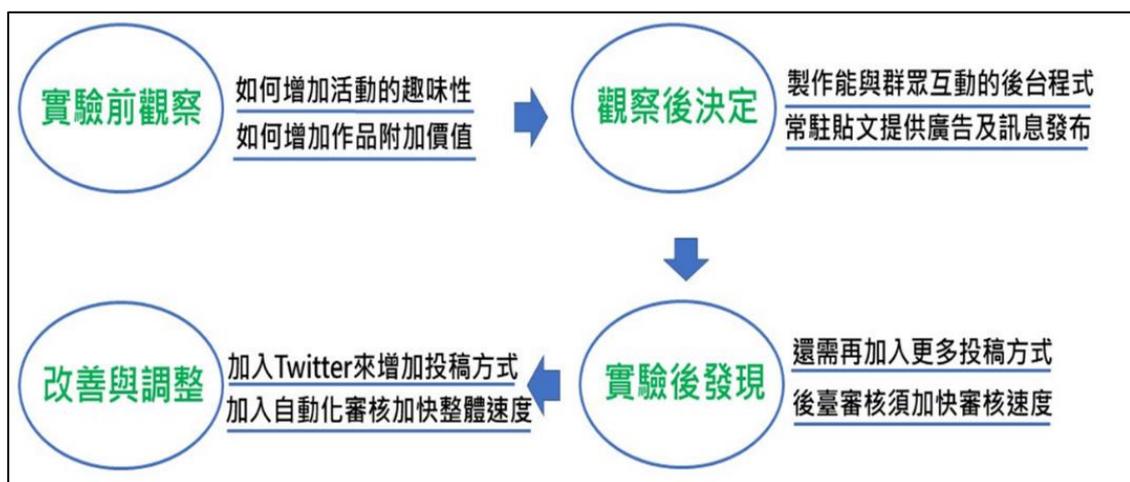


【圖二十四】自動審核流程圖

五、實驗與問卷調查

我們在學校的校慶活動進行實驗，首先我們在活動當中發放問卷，以 Google 表單的方式觀察群眾對於一般大型活動的感覺，並且回收有效問卷進行統計，以此來討論如何增加活動的趣味性與作品附加價值。

觀察後我們決定製作能與群眾互動的貼文推播系統，以及加入常駐貼文進行廣告投放，增加商業價值。完成後，我們將作品投入至校慶活動當中，再次進行問卷調查，得知尚未發現的問題後，將觀眾回饋的問題加以修正，精進整套系統，如【圖二十五】所示為實驗流程圖。



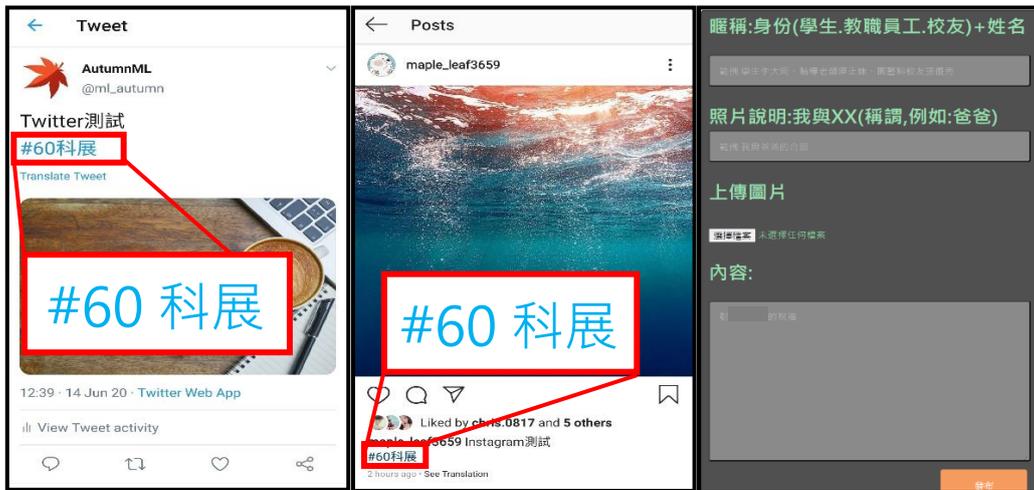
【圖二十五】實驗流程圖

伍、研究結果

一、系統功能

(一)利用社交軟體或網頁平臺投稿

這套系統使用了最受年輕人歡迎的社群軟體 Instagram 和國際化的 Twitter 作為投稿平臺，為了讓和上述平臺不熟悉的使用者投稿，所以我們也提供了網頁投稿的服務，讓投稿方式更加豐富，同時若是使用社交軟體投稿，只需在自己的頁面發文並加上 hashtag，我們的系統變會自動將此貼文擷取投稿，如【圖二十六】所示。



【圖二十六】各平臺投稿頁面

(二)LINE BOT 自動審核

利用 LINE 群組機器人和網頁作為審核貼文的平臺，以方便多位管理員能夠同步審核貼文，並加入自動審核讓貼文審核得更快，如【圖二十七】、【圖二十八】所示。



【圖二十七】網頁審核頁面



【圖二十八】LINE 審核頁面

(三) 訊息輪播在大螢幕上

審核完的貼文將會進入下一輪的輪播中，讓使用者和大眾分享訊息、感想，不只加快了訊息的流通，更讓活動主辦方可以和群眾互動使活動錦上添花，我們同時限制了輪播貼文的最大數量，讓每個人都有機會看到自己的貼文，不會等了老半天還輪不到，如【圖二十九】、【圖三十】、【圖三十一】所示。



【圖二十九】 Instagram 投稿輪播 【圖三十】 Twitter 投稿輪播 【圖三十一】 網頁投稿輪播

(四) 進行廣告與訊息傳播

除了民眾的投稿以外，系統後臺可以刊登常駐貼文，貼文不會被後面上傳的貼文覆蓋，可以提供給主辦單位或廠商進行消息的公布或是工商廣告，如【圖三十二】所示。



【圖三十二】 常駐貼文輪播

(五) 系統模組化方便主辦方設定主題

把背景、標題、hashtag……等眾多自訂項目整合到了程式裡，方便活動單位設

定各項主題應用於不同的場合和活動中，如【圖三十三】、【圖三十四】所示。



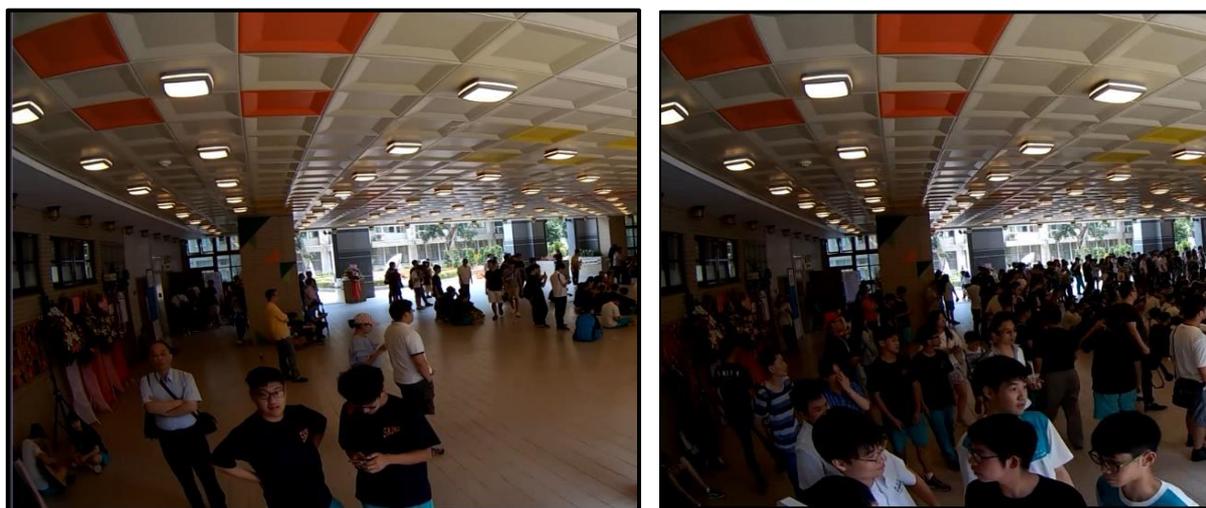
【圖三十三】登入介面



【圖三十四】管理介面

(六)達到增加活動趣味性的目的和宣傳效果

總結上述所有功能，加上和主持人與活動的連動，搭配良好的位置與大螢幕，此系統可以有效地大幅增加活動的趣味性及互動性，如【圖三十五】所示。



【圖三十五】師生參與活動和利用系統

二、審核速度比較

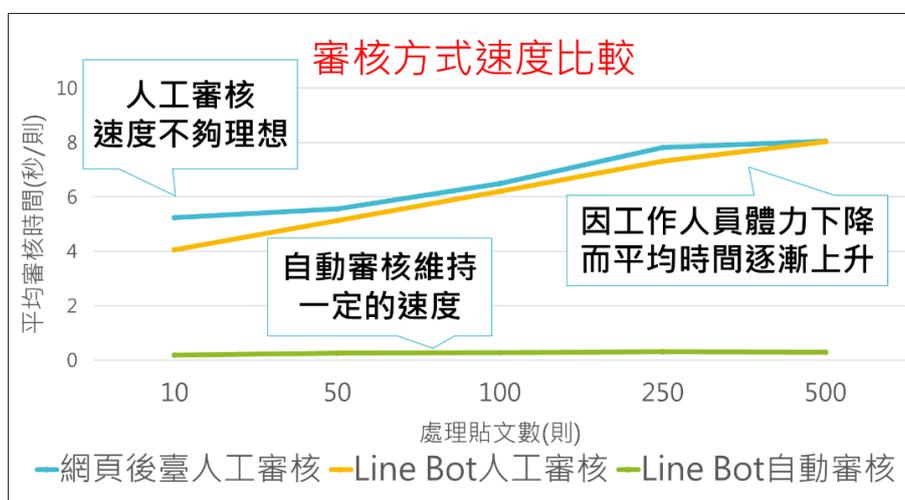
我們對網頁後台人工審核、Line Bot 人工審核、Line Bot 自動審核進行速度比較，以相同貼文量觀察每則貼文審核所耗時間，發現以網頁後台人工審核方式進行審核不僅耗

費大量人力，且速度不盡理想，後來我們改採用 Line Bot 讓管理員不必守在電腦前也可審核貼文。從【表五】與【圖三十六】可知，雖然在貼文數少的時候 Line Bot 人工審核速度稍有增快，但在貼文量多時會因工作人員體力下降而導致平均時間上升。我們最後結合圖像辨識系統來讓 Line Bot 自動審核貼文，可從【表五】與【圖三十六】看出自動審核可維持一定速度，適合大型活動採用。

【表五】審核方式速度比較表

審核方式 貼文量	網頁後臺 人工審核	Line Bot 人工審核	Line Bot 自動審核
10則	5.23秒/則	4.05秒/則	0.17秒/則
50則	5.55秒/則	5.12秒/則	0.25秒/則
100則	6.49秒/則	6.20秒/則	0.26秒/則
250則	7.80秒/則	7.31秒/則	0.30秒/則
500則	8.04秒/則	8.02秒/則	0.29秒/則

(表五資料來源：自行設計)



【圖三十六】審核方式速度比較折線圖

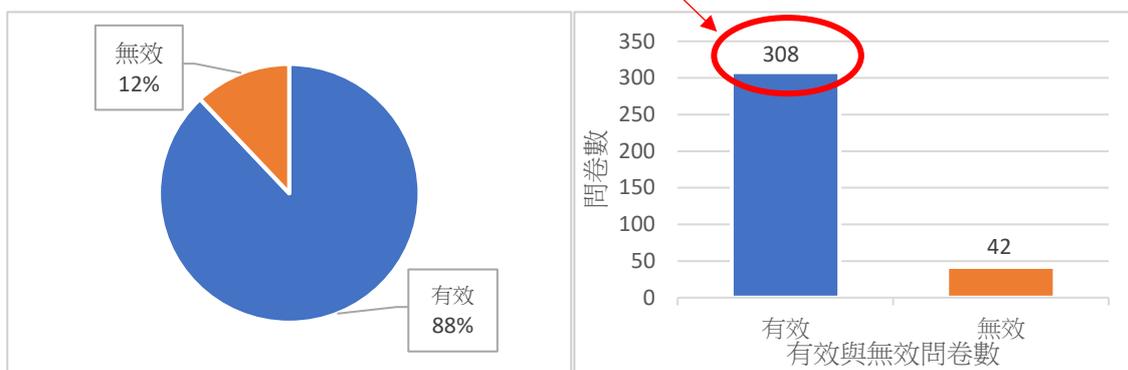
三、問卷調查結果

由於先前學校活動當中並沒有使用這項作品，因此我們在學校校慶時進行我們作品的測試，並使用 Google 表單進行問卷調查，希望藉由活動驗證我們對於群眾使用社群軟體的看法和收集大家的看法以及確認這項作品是否能夠令活動更加有趣及熱絡。

(一)使用本作品前

我們在學校的校慶表演的問卷調查約 350 份，回收後的問卷經過確認，有效問卷約有 308 份，占總發放問卷的 88%，用以確認此次校慶大眾的感覺如何，如【圖三十七】所示。

使用作品前發放的有效問卷為 308 份

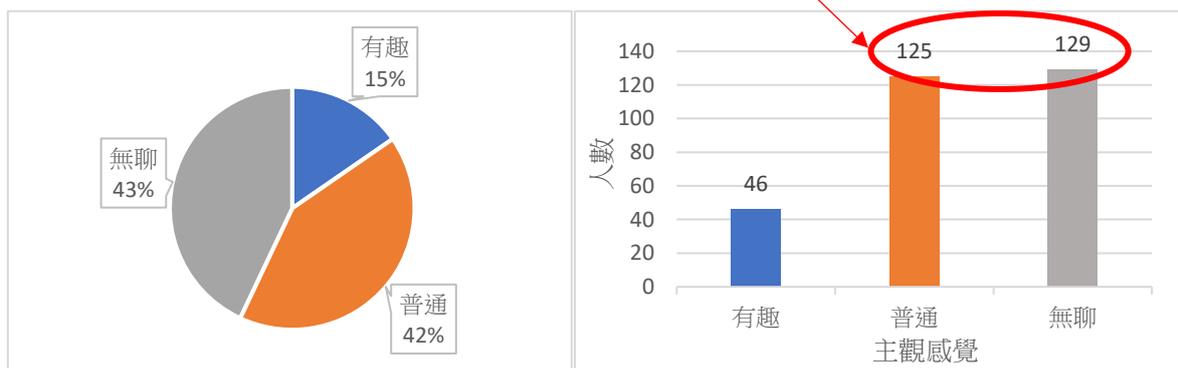


【圖三十七】使用作品前發放的有效問卷占比

1.使用者對活動的整體感覺

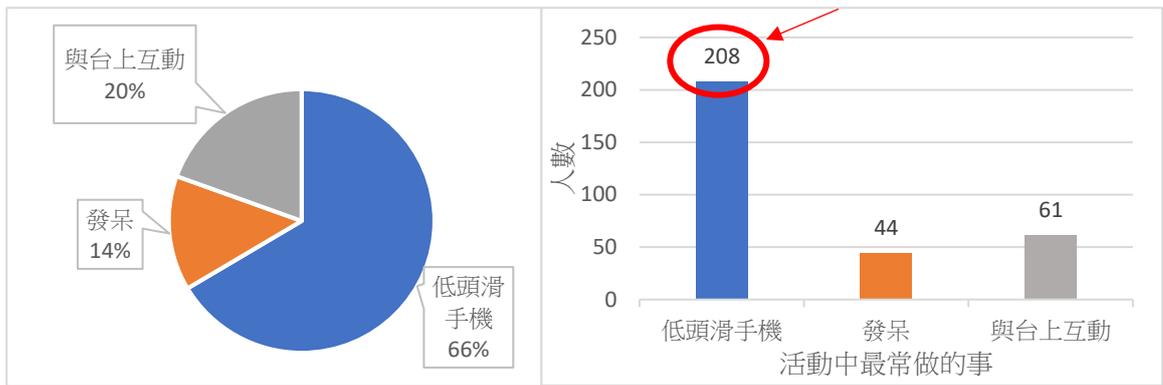
我們於問卷中詢問大家對於活動中的整體感覺如何，且詢問他們在活動當中花費大多數時間在做什麼事情，如【圖三十八】、【圖三十九】所示。

群眾對於以往活動較不滿意



【圖三十八】大眾對於活動的整體感覺

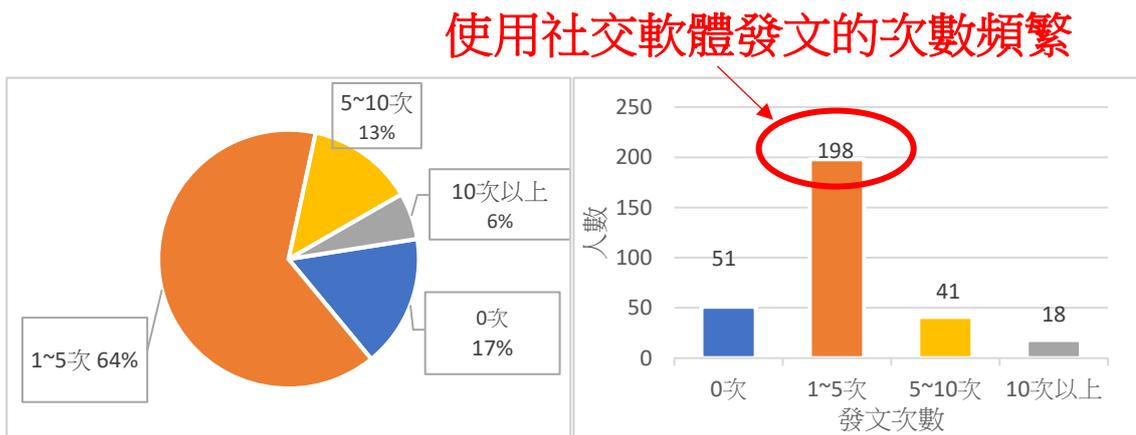
大眾太少融入活動



【圖三十九】大眾在活動中在做什麼事情

2. 在活動中使用社交軟體的平均發文次數

我們在問卷中詢問學生在活動當中使用社交軟體的平均發文次數，使我們發現使用社交軟體的人時常發文以表達心情，且次數大約有 1~5 次或以上，如【圖四十】所示。

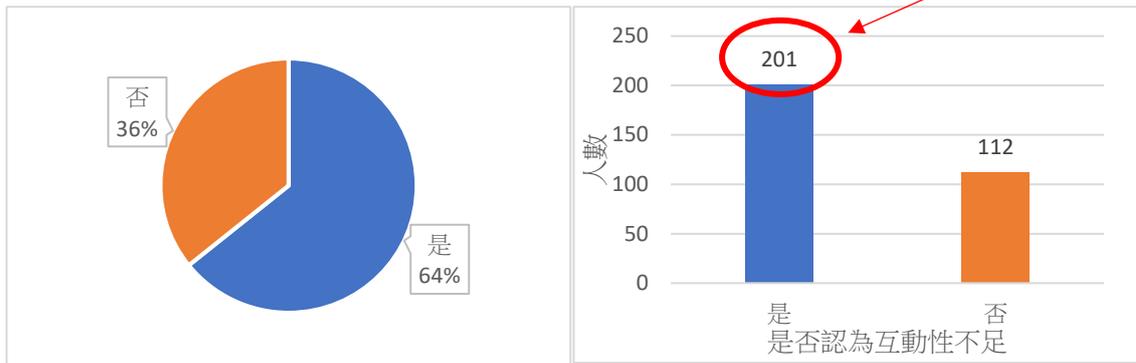


【圖四十】在活動中使用社交軟體的平均發文次數

3. 是否認為臺上與臺下互動不足

我們在問卷中詢問他們是否認為活動中臺上與臺下的互動不足，大部分的回答都給予了回饋，表示目前活動中應增加臺上與臺下之間的互動，如【圖四十一】所示。

大多數人認為互動不足



【圖四十一】是否認為臺上與臺下的互動不足

4.總結

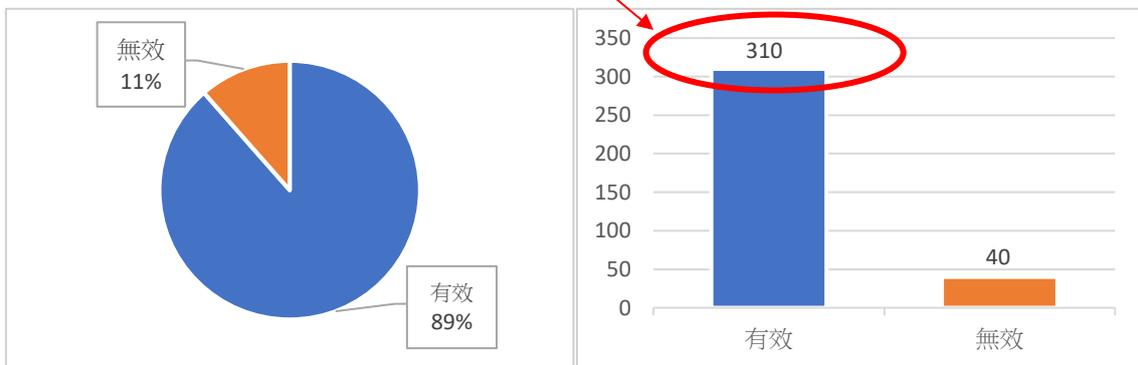
由上述的問卷調查中我們可以發現：高達 43%的人覺得活動很無聊，表示目前大多數的活動都急需改變，而另一方面，有高達 66%的人在滑手機，其中更有 63%的人使用社交軟體超過 2~3 小時以上，導致臺上與臺下的互動性嚴重不足，致使活動變得尷尬、冷場。

(二)使用本作品後

我們在問卷的第二階段詢問大家對於使用作品後對活動的感覺，收集一些意見好看出今年活動是否有所不同，並收集一些使用者的使用習慣，以利用於未來改善整體作品。我們收集了 350 份問卷，其中有效問卷為 310 份，約佔總數的 89%，如

【圖四十二】所示。

使用作品後發放的有效問卷為 310 份

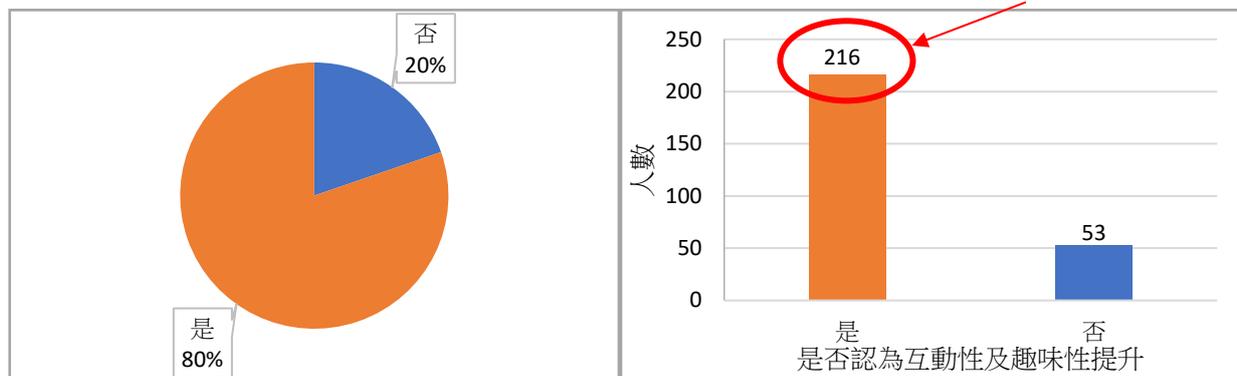


【圖四十二】使用作品後發放的有效問卷占比

1.對於增加活動的趣味性與互動性

我們向使用者詢問他們對於這套系統和活動之間的連動是否有使活動更加有趣，大部分的使用者給予了正面回饋，認為整體互動及趣味性提升的人高達 80%，表示這套系統讓他們與臺上的人擁有更多互動，如【圖四十三】所示。

作品確實提升了活動的互動及樂趣

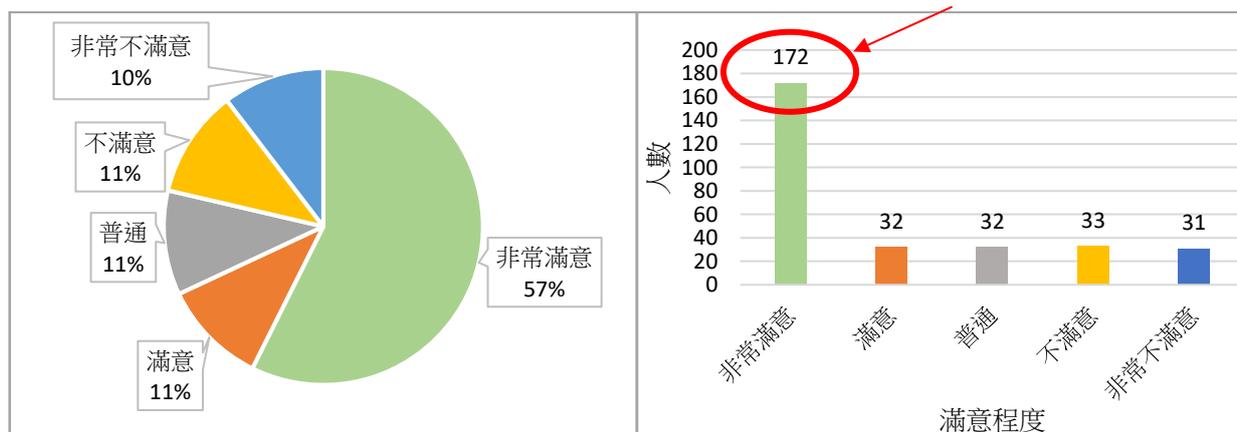


【圖四十三】 是否有提升活動互動性和增加活動的趣味性

2.對於系統和活動連動的滿意度

我們向使用者詢問這次活動和系統的連動滿意度，雖然 57%的人持正向看法，但看得出來我們在連動性這塊還有進步空間，我們需要較多配套和方法讓系統更像活動的一部分，如【圖四十四】所示。

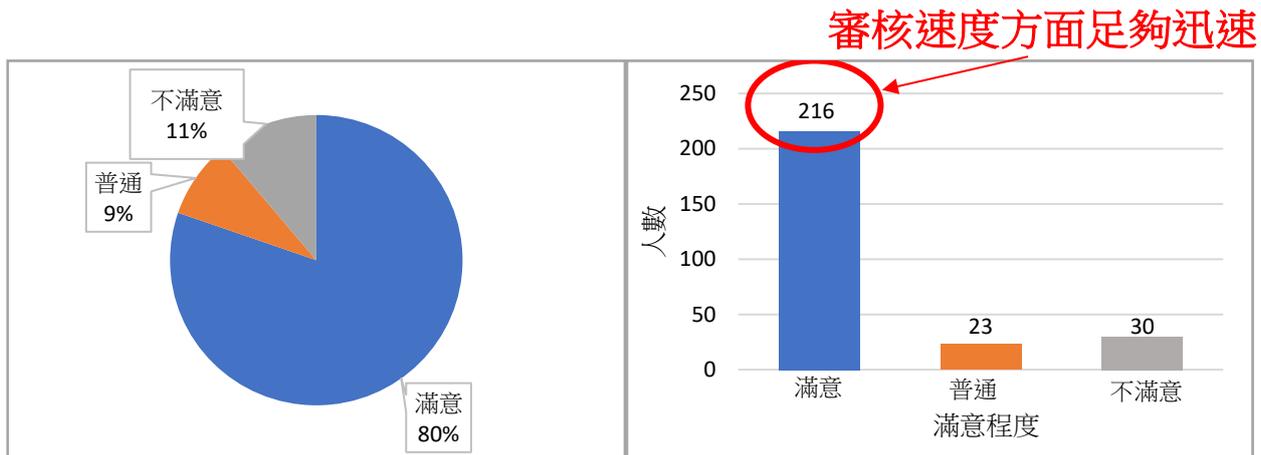
作品在活動中的連動性足夠



【圖四十四】 對於系統和活動連動的滿意度

3.對於審核速度的滿意度

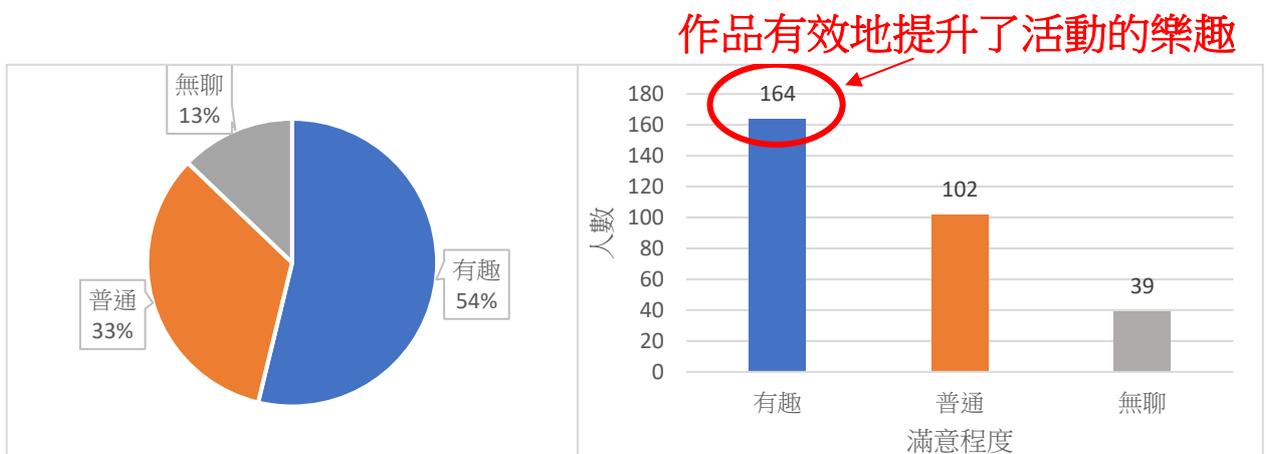
我們實驗時用在校慶的小型活動中採用人工審核，並且因為我們有三個人，因此滿意度非常高，有 80%的人覺得非常滿意，但是我們認為如果更大型的活動中有限的人力可能拖慢審核貼文的速度，於是我們加入了自動審核，以解決在大活動中審核過慢的問題，如【圖四十五】所示。



【圖四十五】對於審核速度的滿意度

4.大眾使用後對於活動的整體感覺

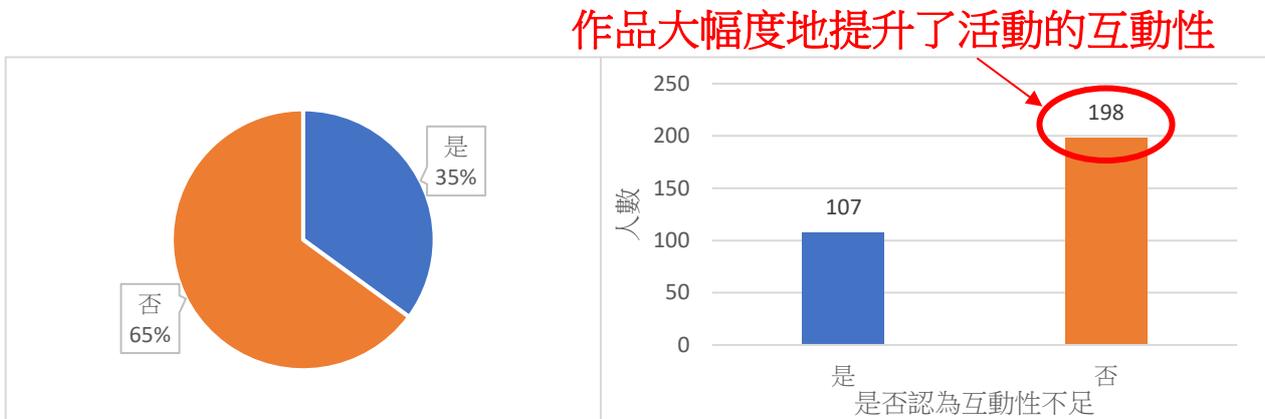
我們在問卷中再次詢問對於活動的整體感覺以及活動當中在做什麼，從圖中可以看見我們的作品有效地提升了活動整體的趣味性，認為活動無聊的人數也減少 30%，而與臺上互動的人數更提升了 40%。如【圖四十六】所示。



【圖四十六】大眾使用後對於活動整體的感覺

5.是否認為臺上與臺下的互動性不足

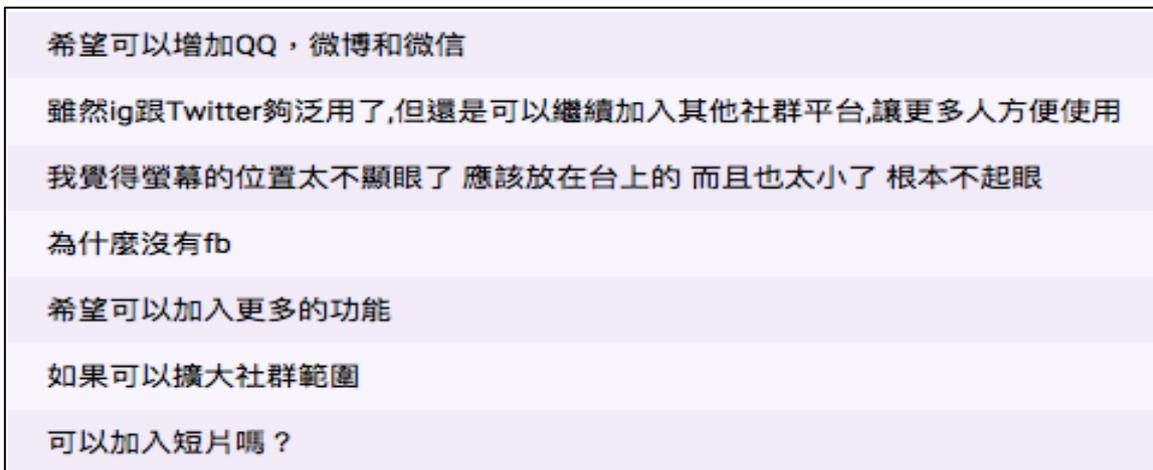
從圖表中我們也能看出在利用此作品後大家對於活動整體的熱絡感有明顯的改善，認為互動性足夠的人數提升 29%，如【圖四十七】



【圖四十七】 是否覺得臺上與臺下互動不足

6.使用者回饋

如【圖四十八】為使用者使用我們的作品之後所給予的回饋，幫助我們進一步提升作品的完整度。



【圖四十八】 使用者給予的回饋

7.總結

雖然我們的作品仍然有些美中不足，像是：螢幕大小及位置等等的要素，不過我們能發現使用此研究後，活動當中臺上與臺下的整體互動有大幅度的提升，大家也比較不專注在滑手機，而是更常與臺上作互動，使得活動整體的熱絡度提升。

陸、討論

以前寫 C#時有 VisualStudio 協助我們更加順利撰寫程式，不僅打出關鍵字就可以自動跳出相關函式，還會以中文提示使用者需要填寫哪些類型的參數。但在 PHP 上使用 VisualCode 就沒有那麼方便了，一開始光是要找如何 Debug 就花了好一段時間，後來還加入了 Ajax 技術來達成動態輪播，當初引用 jQuery 來寫 Ajax 時遇到了些許困難，包含 Ajax 丟出資料後無法觸發成功事件(Success)的函數，嘗試了好幾次沒有錯誤(Error)的問題，發送前(BeforeSend)的函數也有正常觸發，由此我們推論資料有送出去但沒有遇到錯誤，經過了一下午的測試和查詢資料，發現是成功事件(Success)函數的名稱拼錯，才導致一直沒觸發，由於是引用 jQuery 撰寫，也不是內建的 JavaScript 函數，所以引用 jQuery 上的程式出錯時並不會主動提醒，研究這方面時花費了許多心力與時間。而我們曾經看到網路上有人有做過抓取 Facebook 貼文的功能，當初我們也想要嘗試抓取 Facebook 的貼文使用在我們的系統中，但一查之下才發現 Facebook 現在不提供全站爬蟲的 API，只能查詢公開的個人塗鴉牆，起因是因為今年 4 月被爆出的劍橋分析事件：2014 年劍橋大學的一位教授因為製作了一款心理測驗 App，被劍橋分析公司濫用來竊取個資並販賣。因此 Facebook 關閉了相關功能以防止類似事件再次發生，也迫使我們暫時放棄這個想法。我們也花了一些時間來研究如何使抓取貼文更加快速，並搭配自動審核來讓貼文得以及時上傳到大螢幕上。查詢文獻後，發現網路爬蟲大致上分為三種策略，其中我們以聚焦搜索策略來實現最快速、有效率的抓取貼文，提升了作品的完整性。

本研究計畫未來的努力方向和工作，包括以下幾點：

- 一、研究有沒有其他替代方案可以讓使用者利用 Facebook 進行投稿。
- 二、套用圖像辨識的 API 來使我們的自動審核更加的完善，能夠自行過濾不當的圖片。
- 三、模組化的方面能夠更加的完善，希望能達到僅僅一個程式就能設定所有相關參數。
- 四、發展適用於大型活動的功能，例如在輪播中插入數則廣告，來幫助活動的贊助廠商進行宣傳，抑或是在輪播中加入主辦方公告的內容。
- 五、將 LINE 機器人功能發展更加完善，希望未來把網頁後台的功能全數移植至 Line 上面。

柒、結論

在這項研究中我們以「利用系統讓活動更有趣」作為主要研究目標而製作「結合社交軟體建置大型活動之影音廣播平臺」，而後衍伸出「在活動中進行商業活動」和「提供主辦方傳遞訊息」的主要功能，且想到活動進行時大多是公開場合，因此必須嚴加控管投稿內容，如此卻又擔心審核過慢，因此想到加入了 LINE BOT 跟網頁後臺審核系統。透過測試及問卷調查，我們進一步的改良系統，加入了自動審核以及模組化系統，更加入了新的社交投稿平臺 Twitter 來擴大不同的使用客層，並且了解我們的系統必須要更大的螢幕以及良好的位置和配套才能發揮百分之百的效益。未來我們將持續的修正問題以及應用在更多的活動中，也利用這些數據提高使用者體驗，使這個研究更趨於完善、穩定。

以下為本研究實驗結果與結論：

一、探討如何讓活動更加有趣

我們在學校活動中發現大部分的人對活動及主持人之間都沒有太多互動，而是自己在臺下聊天或玩手機，因此我們設計了這套系統希望能增加活動與觀眾之間的互動性，並利用問卷來確認是否能讓活動更有趣。

二、探討出席活動的群眾使用不同社交軟體的比例

我們觀察網路上的統計資料發現時下年輕人使用 Instagram 的比例大幅增加，因此我們決定以 Instagram 作為第一個加入系統的社交平臺，之後我們也在學校活動中進行測試，證實我們的想法正確，並且嘗試加入更多的社群平臺進行投稿，可惜的是在臺灣很多人使用的 Facebook 因為一些原因無法進行全站的爬蟲所以目前尚未加入系統。

三、探討使用前與使用後的差異

如研究資料的圖表及數據所示，在使用此系統前僅有 15% 的人認為活動是有趣的，但在使用這套系統之後，認為活動變得有趣的使用者有 80%，且認為整體活動變得有趣的人也提升到了 53%，我們能從這點看出這套系統在活動中能達到大幅的改變，讓活動中的人們更沉浸於活動當中。

四、探討如何讓使用者輕鬆的使用這套系統

我們使用 Instagram、Twitter 和網頁做為平臺，使用者只需在貼文中加入 Hashtag 即可

使自己的貼文被大眾看見，讓大家都輕鬆的使用，藉此提升使用者對於活動的參與感，改善現今大型活動的缺點。

五、探討如何加快審核速度，讓活動與系統即時同步

我們加入自動審核系統，讓貼文能夠更快顯示到系統上，使用者跟活動的互動可以更即時，訊息傳播也可以更快速，希望能讓大家更想要與台上互動。

六、探討使用何種圖像辨識技術最具效率

為使自動審核系統更加快速，我們做了許多文獻探討，發現有邏輯回歸、自適應增強、決策樹三種圖像辨識技術。經討論與實驗後發現「自適應增強」最符合我們研究的審核系統，以最有效率的方式審核觀眾的貼文，節省人力之外更能提升在活動當中的審核速度。

七、比較各種網路爬蟲策略，加快網路爬蟲速度

為了有效抓取使用者發布的貼文，我們在網路上搜尋並找到了三種網路爬蟲策略，且經過比較後發現「聚焦搜索策略」符合我們系統的需求，更快速的抓取資料，以達成及時推播貼文的功效。

八、利用 Line Bot 技術為工作人員打造方便易用的管理工具

本作品利用 Line Bot 的易用性，將貼文內容、審核結果等資訊以 Line 訊息通知管理人員，並且可在 Line 介面中操作手動通過、刪除、封鎖貼文，使工作人員不必守在電腦前操作審核系統，更方便地在大型活動中管理大眾的貼文。

九、探討如何更加有效的利用我們的系統

我們在校園中測試過後發現，此系統必須搭配完善的配套與主持人進行搭配並且放置於顯眼的位置還要有足夠的螢幕大小，達成上述條件便能讓此系統達到最高效的應用，未來也希望能以更低的環境需求達到更高的效益。

十、探討此系統的附加價值

在學校活動中我們發現此系統可以增加更多的功能，例如：訊息公布、工商廣告……等等，因此我們做了常駐貼文功能，提供了迅速的訊息傳播，且提升了此作品商業化的可能性。

捌、參考資料及其他

- 莊文彬(2016)。國際即時通訊軟體使用行為意向研究-以 LINE、FB Messenger 為例。文藻外語大學：碩士論文。
- 石莉平(2013)。行動即時通訊軟體媒介豐富度之互動效果研究—以大學生為例。世新大學：碩士論文。
- 黃佳惠(2012)。即時通訊文字訊息對話特性。國立臺北科技大學：碩士論文。
- 彭威鳴(2014)。即時通訊軟體之消費者行為研究。健行科技大學：碩士論文。
- 葉修毓(2007)。即時通訊使用行為影響因素之初探研究。國立中正大學：碩士論文。
- 管郁君(2004)。探討即時通訊之使用動機、使用行為與滿足之關係。國立政治大學：碩士論文。
- 張華明(2018)。基於網路爬蟲和聲控技術應用於線上音樂點歌之研究。聖約翰科技大學：碩士論文。
- 環博文化(2016)。PHP+MySQL+Dreamweaver 動態網站開發從入門到精通（第二版）。北京市：機械工業出版社。
- 傳智播客高教產品研發部(2016)。SSH 框架整合實戰教程。新竹市：清華大學出版社。
- 鄧文淵、黃信義、文淵閣工作室(2014)。挑戰 PHP/MySQL 程式設計與超強專題特訓班（第三版）。台北市：碁峰資訊股份有限公司。
- 鄧士昌(2015)。Ubuntu 完全自學手冊-桌面、系統與網路應用。新北市：博碩文化股份有限公司。
- 機器學習 – 定義、演算法、以及商業應用。2020 年 6 月 13，取自 <https://oosga.com/machine-learning/>
- CH.Tseng(2017)。圖像特徵比對(一)-取得影像的特徵點。2020 年 6 月 13 日，取自 <https://reurl.cc/j72gEp>

【評語】 052507

此作品應用網路爬蟲技術抓取使用者貼文，設計一套能提高互動性的訊息發送平臺，使用 LineBot 結合圖像辨識自動且快速審核貼文，另外開發一個後台系統，方便活動單位常駐貼文輪播活動訊息、廣告與管理。此作品是一個訊息提取、審查、和發佈的系統，在抓取訊息時採用聚焦搜索策略，在進行自動圖像辨識時使用一些現有的機器學習模型和訓練技術，希望能自動過濾不當或不雅的圖片，通過審查的圖片則傳送到大型螢幕顯示。整體來說，此作品功能算相當完整，但在進行自動圖像辨識希望偵測不當和不雅圖片時，雖然使用一些既有機器學習模型和訓練技術，但沒有深入討論而只是膚淺的帶過，且沒有顯示和討論自動圖像辨識的正確率，因此在這方面未來可以加強改進。

摘要

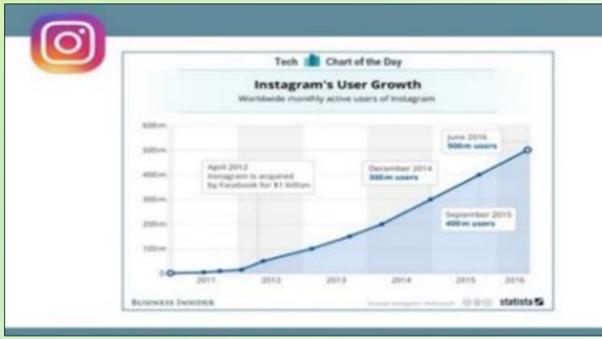
我們發現在校慶或演講活動當中，學生的參與意願不高，時常低頭**使用手機**而對活動整體不感興趣，使活動中的**互動愈來愈冷漠**。因此我們**使用問卷抽查法**探討大眾對於活動的觀點，搭配**網路爬蟲**抓取貼文，觀眾只需在Instagram、Twitter等社交軟體中發布文章，系統就會將文章即時推播至活動中的大螢幕上，並且搭配**圖像辨識**、**關鍵字比對**、**LINE Bot**、**後臺管理**，及**常駐貼文**等功能，便於管理之外也提升商業價值，為各類大型活動量身打造，提供**大幅改善**活動**互動性**與**精采程度**的解決方案。

壹、研究動機

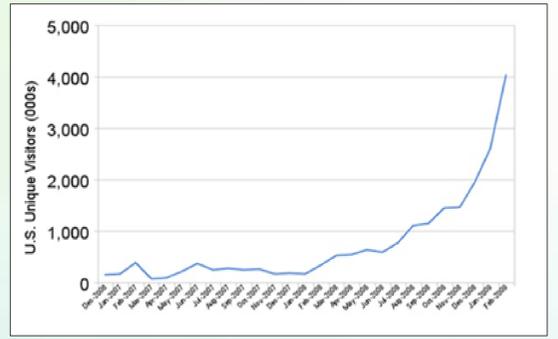
經由問卷調查得知，約有85%的大眾對於校園或節慶活動等興致缺缺，為了改善觀眾的參與感，並增加活動互動性，我們開始研究一套系統，能將現場觀眾的貼文轉貼至大螢幕上，使大眾對於活動的心情得以快速散播，藉由網路爬蟲抓取**和活動最為相關的貼文**，希望能夠**提升群眾參與活動的意願**，且**改善互動性不足的問題**，並為活動**增添看點**。



活動中低頭使用手機



Instagram活躍用戶統計圖



Twitter全球用戶統計圖

貳、研究目的

- 一、研究如何**提升活動的互動性**
- 二、探討不同的**社交軟體使用人數**比例
- 三、探討**網路爬蟲策略**比較
- 四、探討如何利用**圖像辨識**、**關鍵字比對**技術過濾不雅貼文
- 五、探討如何**有效提升審核速度**
- 六、探討使用**互動平台****前後的互動性**比較



系統特色

參、研究過程及方法

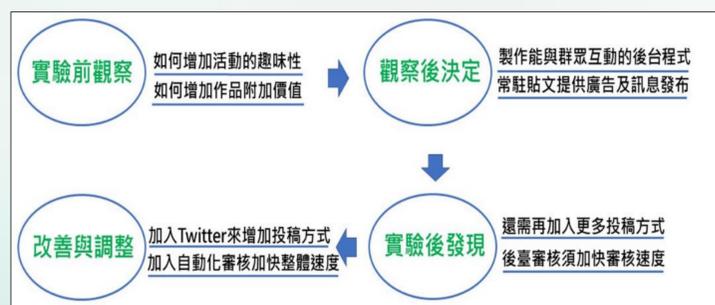
研究設備及軟體

硬體設備	開發軟體	伺服器系統
伺服器主機	PHP	Ubuntu
智慧型手機	Line Bot	MySQL
大型螢幕	Microsoft Azure	SSH Sever

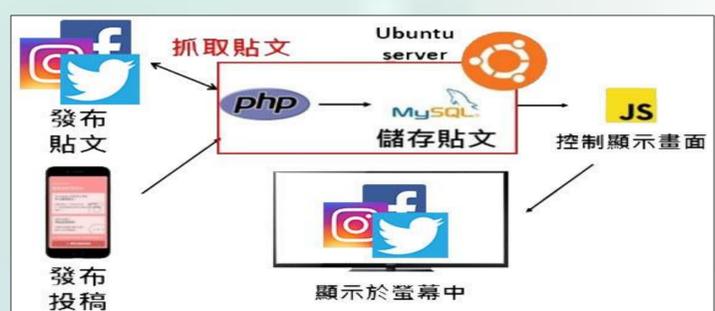
網路爬蟲策略

名稱	方法	概念圖
深度優先策略	網路爬蟲會從起始點開始，一個分支一個分支挖掘資料，進行 深度的搜索 ，直到找到目標資料或是尋遍所有節點與分支。	
寬度優先策略	網路爬蟲會從某一網站開始走訪，接著走訪此網站所有相鄰且未拜訪過的網站，從這些網站進行 先廣後深的搜尋 ，直到找到目的節點或搜尋全部節點。	
聚焦搜索策略	根據「 匹配優先原則 」對數據進行訪問，主動 選擇與主題相關的資料 ，並 限定優先級 ，讓程式可以 快速有效 地獲得更多 與主題相關的頁面 。	

研究流程



系統流程圖



圖像辨識技術

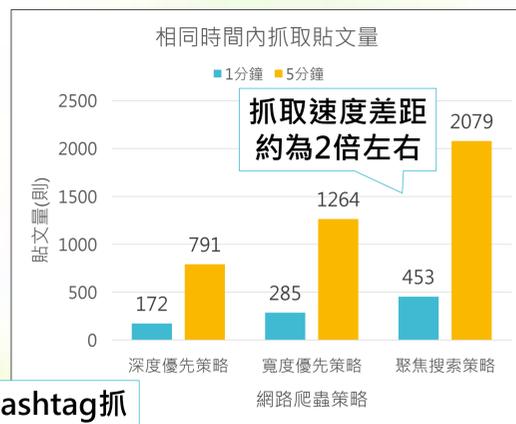
名稱	方法	概念圖
邏輯回歸	最基本的模式是線性迴歸模式，線性迴歸是假設應變數的各個數值是自變數所構成的某種 直線函數值 ，再加上一個 誤差值 所得到的數值，應用於 可視化分析 。	
自適應增強	自適應增強整合多種不同的模型來做分類或回歸分析，再根據其正確率去判斷每一個不同的輸出所佔的比重為何。基於自適應增強的原理，能設計出 特徵點配對 ，應用於 低成本 的方式來做 圖片識別 。	
卷積神經網路	將圖片分割成不同區塊，在由 點的比對 轉成 局部的比對 ，透過一塊塊的 特徵研判 ，逐步堆疊綜合比對結果，如此一來就能 提高判斷準確率 及 判斷速度 。	

肆、研究結果

1. 網路爬蟲策略比較(以相同時間觀察總抓取貼文量)

比較	深度優先策略	寬度優先策略	聚焦搜索策略
1分鐘	172則	285則	453則
5分鐘	791則	1264則	2079則
以主題進行搜索	否	否	是

勝

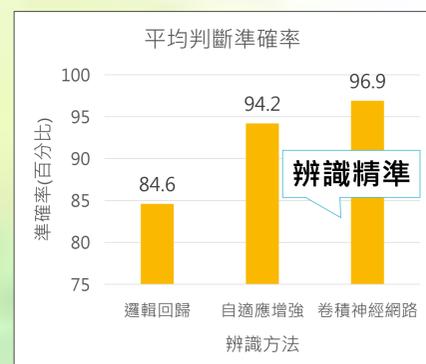
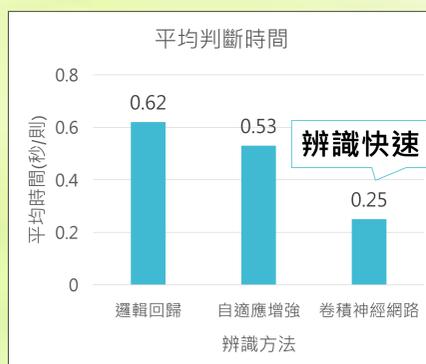


可利用hashtag抓取最為相關的貼文

2. 圖像辨識技術比較(在貼文量為1000則的狀況下進行比較)

比較	邏輯回歸	自適應增強	卷積神經網路
平均判斷時間	0.62秒/則	0.53秒/則	0.25秒/則
平均判斷準確率	84.6%	94.2%	96.9%
圖片特徵點判斷	否	是	是

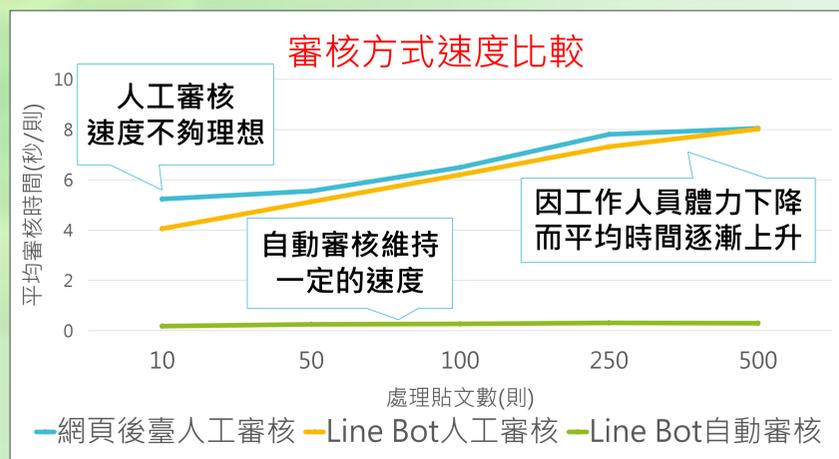
勝



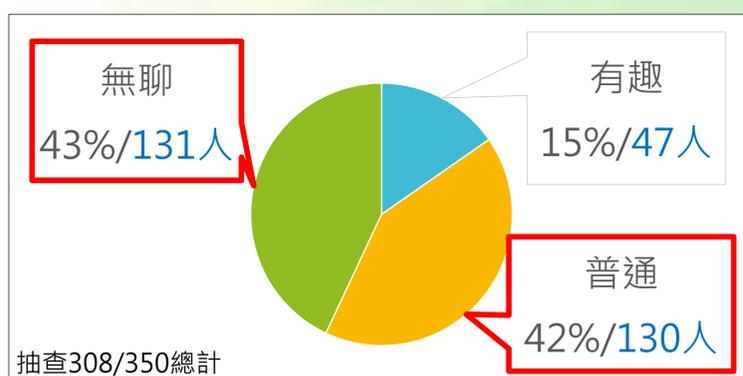
3. 審核方式速度比較(以相同貼文量觀察每則貼文審核所耗時間)

貼文量	網頁後臺人工審核	Line Bot人工審核	Line Bot自動審核
10則	5.23秒/則	4.05秒/則	0.17秒/則
50則	5.55秒/則	5.12秒/則	0.25秒/則
100則	6.49秒/則	6.20秒/則	0.26秒/則
250則	7.80秒/則	7.31秒/則	0.30秒/則
500則	8.04秒/則	8.02秒/則	0.29秒/則

勝

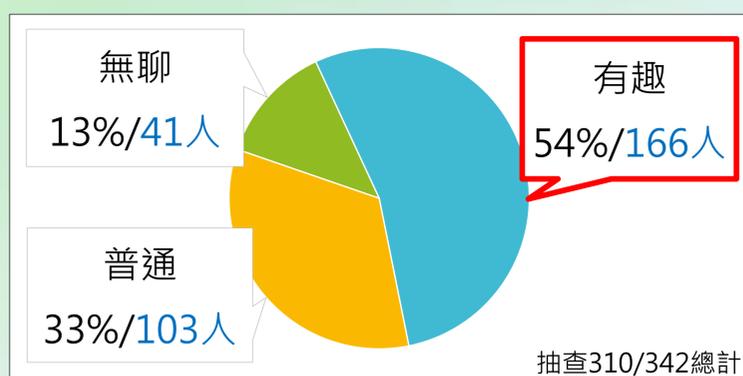


4. 使用者對活動的整體感覺比較(採用隨機抽查10個班的方式：使用前發放350份問卷，使用後發放342份問卷)



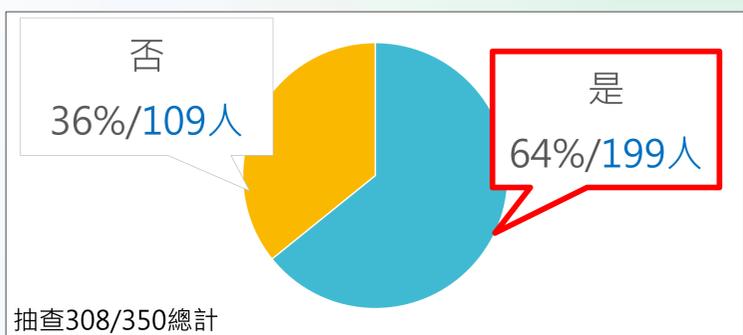
(實驗本研究前 群眾對於以往活動較不滿意)

將本研究投入至活動當中



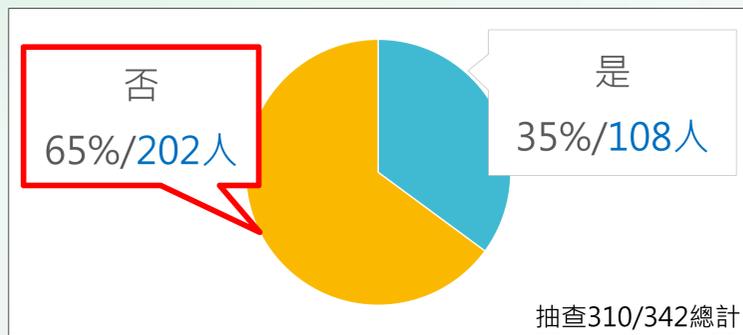
(實驗本研究後 作品有效地提升了活動的樂趣)

5. 臺上與臺下的互動性是否不足(採用隨機抽查10個班的方式：使用前發放350份問卷，使用後發放342份問卷)



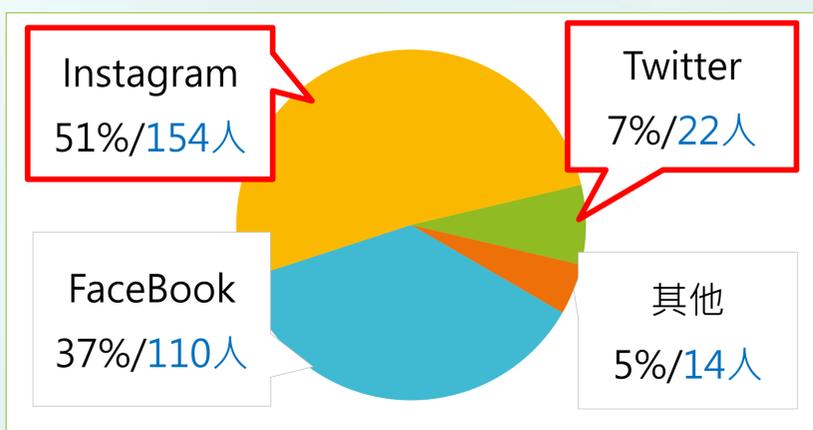
(實驗本研究前 大多數人認為互動不足)

將本研究投入至活動當中



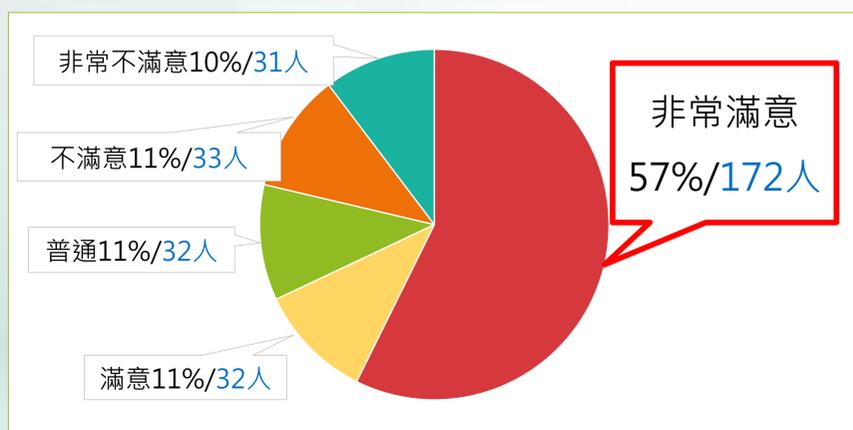
(實驗本研究後 作品大幅度地提升了活動的互動性)

6. 民眾使用社交軟體比例



(民眾使用的社群軟體以IG為主)

7. 民眾對於審核速度的滿意度



(作品在活動中的連動性足夠)

伍、討論

一、是否有其他方法提升活動互動性？

我們透過社交網站的特性，製作出能與群眾互動的系統，讓使用者更融入活動整體，使活動整體更加具有互動性。

二、如何利用網路爬蟲擷取社交網站資料，並更快速的抓取貼文？

我們目前以「**聚焦搜索策略**」技術，使網路爬蟲**更快速、更有效率**的抓取貼文。

三、探討群眾使用不同社交軟體的比例

我們從問卷調查中發現時下年輕人**使用Instagram的比例最多**，其次則是Facebook，而Twitter使用率最低。

四、探討此系統未來的附加價值

在校園中幾次的實驗後我們決定加入「**常駐貼文**」系統，可進行廣告宣傳或訊息發布，提升研究**商業化的可能性**。

五、如何使審核速度更加快速，並更精準的過濾貼文？

我們的**自動審核系統**，目前是以「**卷積神經網路**」與「**關鍵字比對**」技術，使系統在大型活動當中**更加精準、快速判斷觀眾貼文**。

六、探討使用前與使用後的差異，並了解民眾需求

經調查僅有15%的人認為以往活動是有趣的，但在使用這套系統之後，**認為改善了活動參與感的使用者有約80%**，確實達到讓觀眾**更沉浸於活動中的目標**。

陸、結論

- 一、我們利用「**網路爬蟲**」抓取觀眾的社交軟體貼文，並同步顯示於現場的大螢幕上，增加活動與觀眾之間的**互動性**，達到**吸引觀眾目光**的效果。
- 二、我們使用經統計得知當今最常使用的Instagram、Twitter作為投稿平臺，使觀眾能以**多樣化及簡單**的方式發布心情到螢幕上，增加**系統可用性**。
- 三、為有效抓取使用者發布的貼文，經過比較後發現「**聚焦搜索策略**」符合我們系統的需求，以達成**及時推播貼文**的功效。
- 四、為了精準並快速過濾使用者發布的貼文，經過比較後得知「**卷積神經網路**」演算法符合活動需求，達到**快速自動審核**的目的。
- 五、提供主辦方「**LINE Bot自動審核**」以及「**系統模組化**」功能來方便自訂各種主題，讓工作人員在大型活動中方便管理。
- 六、經由問卷調查得知，在活動中使用我們的系統之後，**確實達到改善活動互動性**的目標。

柒、未來展望

- 一、我們希望能**尋找其他替代方法來抓取Facebook貼文**。
- 二、分析觀眾的貼文資料，推播更具**吸引力及相關性的廣告內容**。
- 三、**增強系統的資安防護**，防止攻擊者癱瘓系統影響活動進行。

捌、參考資料及其他

- 莊文彬(2016)。國際即時通訊軟體使用行為意向研究-以LINE、FB Messenger為例。文藻外語大學：碩士論文。
- 石莉平(2013)。行動即時通訊軟體媒介豐富度之互動效果研究 - 以大學生為例。世新大學：碩士論文。
- 黃佳惠(2012)。即時通訊文字訊息對話特性。國立臺北科技大學：碩士論文。
- 彭威鳴(2014)。即時通訊軟體之消費者行為研究。健行科技大學：碩士論文。
- 葉修毓(2007)。即時通訊使用行為影響因素之初探研究。國立中正大學：碩士論文。
- 管郁君(2004)。探討即時通訊之使用動機、使用行為與滿足之關係。國立政治大學：碩士論文。
- 張華明(2018)。基於網路爬蟲和聲控技術應用於線上音樂點歌之研究。聖約翰科技大學：碩士論文。
- 環博文化(2016)。PHP+MySQL+Dreamweaver動態網站開發從入門到精通 (第二版)。北京市：機械工業出版社。
- 傳智播客高教產品研發部(2016)。SSH框架整合實戰教程。新竹市：清華大學出版社。
- 鄧文淵、黃信義、文淵閣工作室(2014)。挑戰PHP / MySQL程式設計與超強專題特訓班 (第三版)。台北市：碁峰資訊股份有限公司。
- 鄧士昌(2015)。Ubuntu完全自學手冊-桌面、系統與網路應用。新北市：博碩文化股份有限公司。
- 機器學習-定義、演算法、以及商業應用。2020年6月13日，取自 <https://oosga.com/machine-learning/>
- CH.Tseng(2017)。圖像特徵比對(一)-取得影像的特徵點。2020年6月13日，取自 <https://reurl.cc/j72gEp>