

東吳大學

巨量資料管理學院學士專題成果報告書

專題名稱：YouTube 留言蜚語 2.0

06170101 高子淇

06170145 周欣德

指導老師：吳政隆老師

中華民國 110 年 06 月

摘要

由於現代大多數人由電視轉向網路平台，在過去大家回到家後會優先打開電視，而如今逐漸轉變成人手一機，使用手機或電腦觀看影片，有鑒於傳統媒體的衰落，新媒體的迅速興起和快速發展，近 93% 的民眾曾造訪過 YouTube。YouTube 逐漸成為現代人生活娛樂的重心，藉此本專題想深入研究民眾對於 YouTuber 的觀感、影響力及形象。YouTuber 目前僅能從觀看瀏覽紀錄或使用者行為等數據來追蹤其影片成效，而缺乏最直接表現使用者情緒的留言分析，因此想幫助 YouTuber 更加了解其觀眾對於他影片的感受，以影片底下的留言為依據，進行輿情分析。

本專題使用作者先前做過的相關專題進行延伸，以過去製作網頁的服務為基底再行優化。我們選擇開發 chrome extensions 套件，讓使用者服務可以更加便利，以 javascript + CSS + html 三項元素來組成套件，當使用者在 youtube 影片頁面時，只需要點擊右上角擴充功能上我們的 logo 圖示，自動會進行爬蟲以及 FastText 模型進行預測，結果便會在留言上自動產生自行設計的五大指標圖示，以圖示代表此則留言的影片喜好、YouTuber 喜好、激動程度、是否具有腥羶色和諷刺。

本專題的實現目的，近程為提供使用者服務，可根據所欲分析的資料，動態蒐集並回饋結果，對於 youtube 的頻道經營者或是個別粉絲都大有幫助。遠程可作為 youtuber 經營頻道的績效指標，並同時讓社會大眾審視目前社會的概況及趨勢。

目錄

摘要	2
目錄	3
圖目錄	4
表目錄	5
緒論	6
介紹背景	6
動機	7
目的	8
專有名詞定義	8
研究步驟	9
研究方法	9
結果呈現	15
結論	17
附錄	18
參考文獻	18

圖目錄

圖 1-1 昨日有看電視之比例	6
圖 1-2 消費者最近一個月曾使用過影視 APP 之比例	6
圖 1-3 台灣用戶使用瀏覽器比例折線圖	7
圖 1-4 全球用戶使用瀏覽器比例折線圖	7
圖 2-1 流程架構圖	11
圖 2-2 模型流程圖	12
圖 2-3 模型流程圖	12
圖 3-1 擴充功能安裝頁面	15
圖 3-2 擴充功能安裝完成後會出現在右上列表	15
圖 3-3 進入 Youtube 影片頁面，點擊右上角 logo 圖示即可使用	16
圖 3-4 分析成果就會呈現在留言上面	16

表目錄

表 2-1 諷刺程度與腥羶色定義	9
表 2-2 影片喜好與 youtuber 喜好定義	9
表 2-3 激動程度定義	10
表 2-4 諷刺程度與腥羶色圖示	13
表 2-5 激動程度圖示	13
表 2-6 影片喜好與 youtuber 喜好圖示	14

緒論

介紹背景

由於新興媒體的開放且可以無所不在和不受約束地表達，改變了傳媒業的秩序，也改變了全球人類的生活方式。相較於傳統媒體，新興媒體的發展更私人化、普及化、自主化，自媒體也應運而生。尤其以透過網路媒體作為發布管道者為甚，加上人手一機和攝影器材的普及，人人都可以拍攝影音資訊上傳到網路平台。根據東方線上 E-ICP 行銷資料庫自 2014 年至 2018 年的資料顯示（圖 1-1），台灣消費者看電視比例逐漸下降。特別是 20-34 歲消費者最為明顯，在 2014 年時仍有高達 8 成 9 有看電視，然而至 2018 年時其收看比例已降至 7 成 9，跌幅高達 1 成。另外在網路影視 APP 上最主要使用族群仍是 20-34 歲（圖 1-2），但 50-64 歲年長者才是使用網路影視 APP 成長最為快速的族群，其使用率五年間成長將近 2.7 倍。由此可見，消費者看電視之比例逐漸下降，網路影視 APP 使用率大幅上升，代表其收視習性已逐漸轉型。

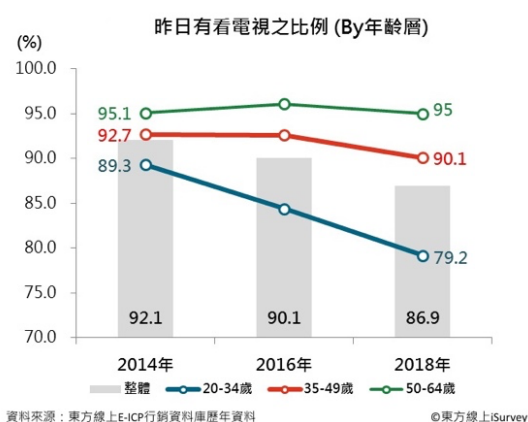


圖 1-1 昨日有看電視之比例

資料來源：許愷洋（2019）。跨代收視習慣大不同，長者網路收視率大漲。iSURVEY 東方線上。取自 <https://www.smartm.com.tw/article/36303632cea3>

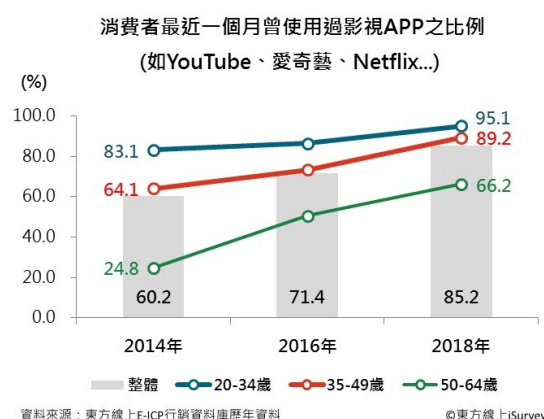


圖 1-2 消費者最近一個月曾使用過影視 APP 之比例

資料來源：許愷洋（2019）。跨代收視習慣大不同，長者網路收視率大漲。iSURVEY 東方線上。取自 <https://www.smartm.com.tw/article/36303632cea3>

動機

根據 Statcounter globalStats 調查，在今年 4 月全球網頁瀏覽器市占率，google chrome 有將近 65% 市占率（圖 1-3），在台灣使用 chrome 瀏覽器的人數也高達 6 成 1 的比例（圖 1-4）。因為有龐大的用戶選擇使用 chrome 來當瀏覽器，許多開發者推出擴充套件來增加應用，提高方便性。另外在 Ipsos 益普索市場研究，台灣每日使用 YouTube 人數高達 7 成，有超過一半的使用者每天平均使用 YouTube 超過 1.5 小時，因此累積龐大的瀏覽量。其中有三分之一的民眾會在購物旅程中使用 YouTube，當中更有超過 8 成民眾表示 YouTube 會影響他們最後購物的決定。由此可見，影音社群平台成為推動品牌策略的得力助手。目前 YouTuber 僅能從觀看次數、觀看時間、點擊次數等使用者行為數據來追蹤其影片成效，而缺乏最直接表現使用者情緒的留言分析，我們先前已完成相關模型，可以來做分析。為了提升使用者服務與方便性，本專題開發外掛套件讓用戶在 chrome 瀏覽器上可以邊觀看影片同時就可以得到留言的輿情分析結果，不需要像過往一樣要到另外的頁面才能得到結果。

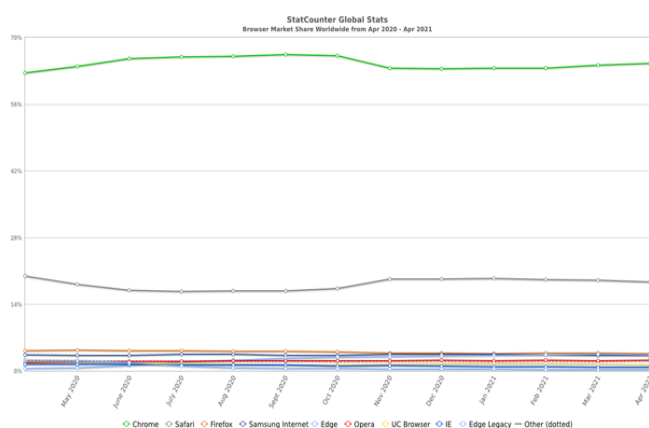


圖 1-3 全球用戶使用瀏覽器比例折線圖

資料來源：StatCounter Global Stats
Browser Market share Taiwan from Apr 2020~Apr 2021。StatCounter。取自
<https://gs.statcounter.com/browser-market-share/all/taiwan>

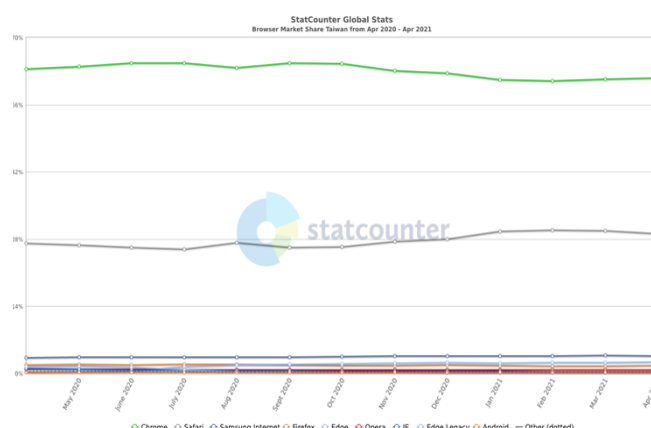


圖 1-4 台灣用戶使用瀏覽器比例折線圖

資料來源：StatCounter Global Stats
Browser Market share worldwide from Apr 2020~Apr 2021。StatCounter。取自
<https://gs.statcounter.com/browser-market-share>

目的

先前我們已完成 YouTube 留言分析的功能，但僅能在我們網站呈現結果，但這樣會增加使用者的不便性，降低使用率。因此我們本次專題以使用者為出發點，希望帶給用戶更好操作體驗，開發 chrome 套件讓使用者可以直接在 youtube 頁面得到分析結果。讓每個 youtuber 在檢視每則留言時，就可以順便了解以下五個指標，對 YouTuber 的喜好程度、對影片的喜愛程度、留言內容激動程度、留言內容諷刺程度、留言內容腥羶色程度，希望從留言的評論中得到觀眾所重視的淺在因素，YouTuber 並可根據不同面向進行改善，拍出符合市場需求的影片。

專有名詞定義

YouTuber：以影音網站 YouTube 為主要活動據點的網路紅人或在 YouTube 投稿之影片創作者。

自媒體：指一般民眾藉由網路手段，向不特定的大多數人或者特定的單個人傳遞規範性及非規範性資訊的新媒體，或稱「草根媒體」、「個人媒體」、「公民媒體」。因部落格、共享協作平台與社群網路（如：臉書、Instagram、微博等）的興起，使每個人都具有媒體、傳媒的功能。

Flask：Flask 為一個使用 Python 撰寫的輕量級 Web 應用程式框架，能提供一個平台來呈現資料。當使用者請求該網址之後，根據函數運算之後回傳給使用者資料，也就是網頁的資料。

Ngork：內網伺服器與外界溝通的一個服務。

Chrome extension:允許工程師採用 Javascript+CSS+html 三項元素，來操作 Chrome 核心功能。

JavaScript：是一種進階、直譯並支援物件導向的程式語言，被世界絕大多數網站使用。

研究步驟

研究方法

1 訂定分數準則

此專題選擇使用高子淇&劉萱&許靖玟&周欣德.YouTube 留言蜚語(Jan , 2021) 進行延伸，使用該專題訓練出的 FastText 模型進行沿用。依照該專題訂定出的留言指標分數如下：激動程度為五種指標，YouTuber、影片內容的喜愛則為三類型的指標判斷，而留言諷刺、腥羶色程度則為二分法。

表 2-1 諷刺程度與腥羶色定義

分數說明	諷刺程度	腥羶色
1	留言中無諷刺語句	留言中無腥羶色字眼
2	留言中有諷刺語句	留言中有腥羶色字眼，甚至攻擊他人的語句

表 2-2 影片喜好與 youtuber 喜好定義

	Youtuber 喜好	影片喜好
1	對 Youtuber 具有負面態度	對影片內容表示負面想法
2	對 Youtuber 態度中立	對該影片內容無表示意見
3	對 Youtuber 具有正向態度	對影片內容有正面肯定

表 2-3 激動程度定義

	1	2	3	4	5
留言內容激動程度	不帶情緒之言論(建議、好像、不太)	些微情緒起伏、哈哈	有情緒起伏、沒想到、哈哈	驚嘆，太、！、真的	感到不可思議，超(級)、千萬、髒話、！！！！

2 架構設計

Chrome extension，採用 JavaScript + CSS+ html 三大元素，來操作核心功能。

2-1 擴充功能的安裝檔(manifest.json)

Chrome extension 套件都擁有一個 JSON 格式的安裝檔，檔名是 manifest，副檔名(.json)，提供了很多重要的資訊，例如這個套件的名字、版本、logo 圖示。並在此檔案中載入會使用到相關資源。如 jquery.js 及 background.js。

2-2 Background.js

在這個檔案當中，我們設置一個點擊功能。當使用者在正確的網址，點擊我們 logo 圖示，便是透過此檔案去啟動我們的功能，去對 content-script.js 下達指令。

2-3 Content-script.js

這個檔案為整個套件主要功能。首先，會透過抓取使用者欲分析的影片 ID 以及當前瀏覽頁面的留言 ID 回傳後端伺服器，來檢查資料是否已經存在資料庫。如果影片 ID 資料未存在的話，便會透過爬蟲程序將相關資料抓取寫入資料庫，再使用模型去進行預測。但如果影片 ID 資料存在的話，會再判斷於使用者前端抓的留言是否存在，如果有未存在之留言，則會按照影片 ID 資料未存在的方式進行。

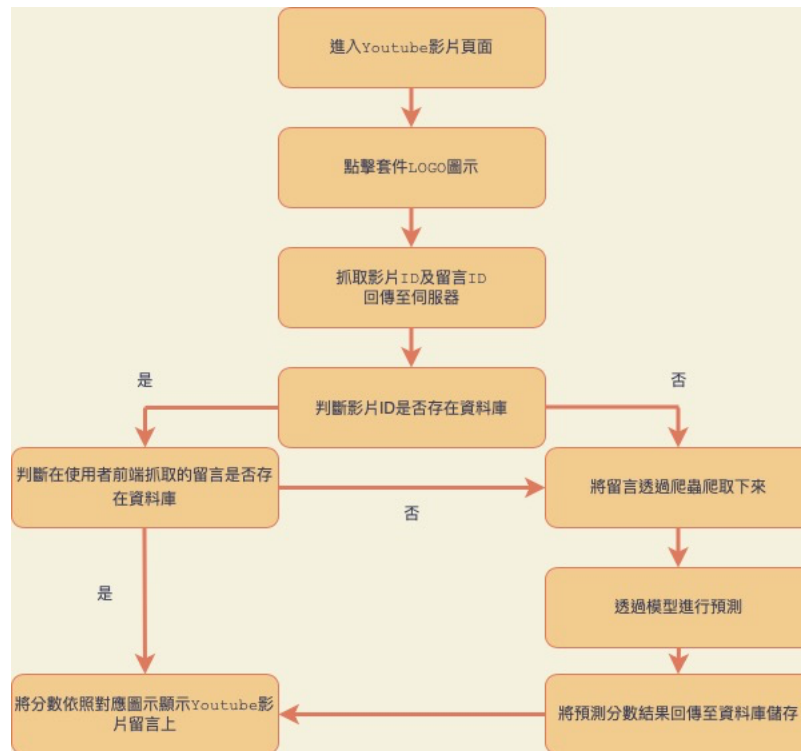


圖 2-1 流程架構圖

2-4 爬蟲程式

參考即會透過參考 Egbert Bouman, (2020, August 7). Re: youtube-comment-downloader. Retrieved from <https://github.com/egbertbouman/youtube-comment-downloader> 抓取 YouTube 影片資訊及全部留言。

2-5 FastText 模型預測

選用高子淇&劉萱&許靖玟&周欣德.YouTube 留言蜚語(Jan, 2021) 訓練出 FastText 該組模型進行留言預測及分析。預測時只需將 CKPT 檔讀取回來，並選擇其對應參數的 py 檔案，即可獲得分析結果。由於不需重複訓練資料，因此大幅節省等待時間。

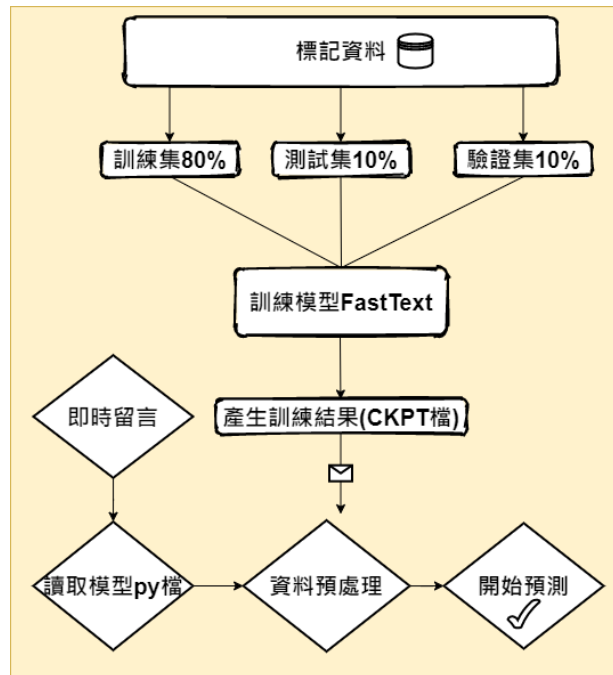


圖 2-2 模型流程圖

3 結果呈現

在高子淇&劉萱&許靖玟&周欣德.YouTube 留言蜚語(Jan , 2021)這篇專題中，各項指標是透過分數來呈現。我們選用該專題的模型作為此次專題使用，因此我們的每項指標一樣會有分數，但當使用者在影片頁面看到這些分數並不一定了解其分數意義，那麼我們不如透過自行設計淺顯易懂的圖示讓使用者知道當則留言背後代表的五項指標為何。下表為我們依照五項指標自行設計的圖示。

表 2-4 諷刺程度與腥羶色圖示


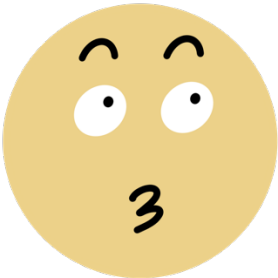
	是	否
是否為腥羶色		無圖示
是否為諷刺		無圖示

表 2-5 激動程度圖示












	1	2	3	4	5
激動程度					

表 2-6 影片喜好與 youtuber 喜好圖示

	不喜歡	中立	喜歡
Youtuber 喜好			
影片喜好			

結果呈現

實際成果

1. 安裝方式：進入 chrome 頁面，選擇「更多工具」裡面的「擴充功能」，進入此頁面點擊「載入未封裝項目」，將我們檔案的資料夾匯入。載入成功，右上角就會出現我們這款套件的 logo 代表安裝完成即可使用。

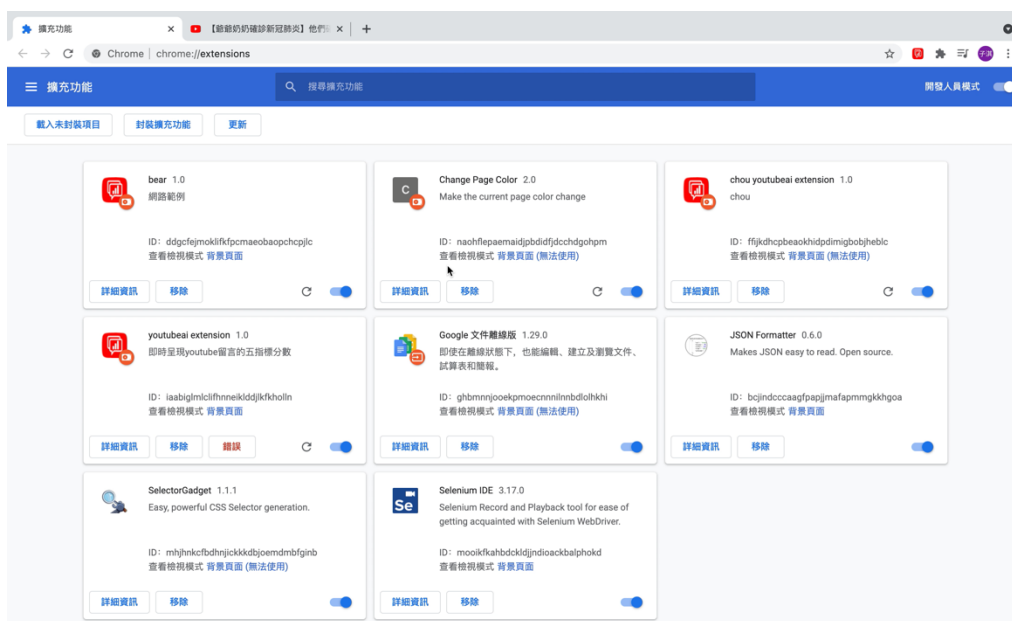


圖 3-1 擴充功能安裝頁面

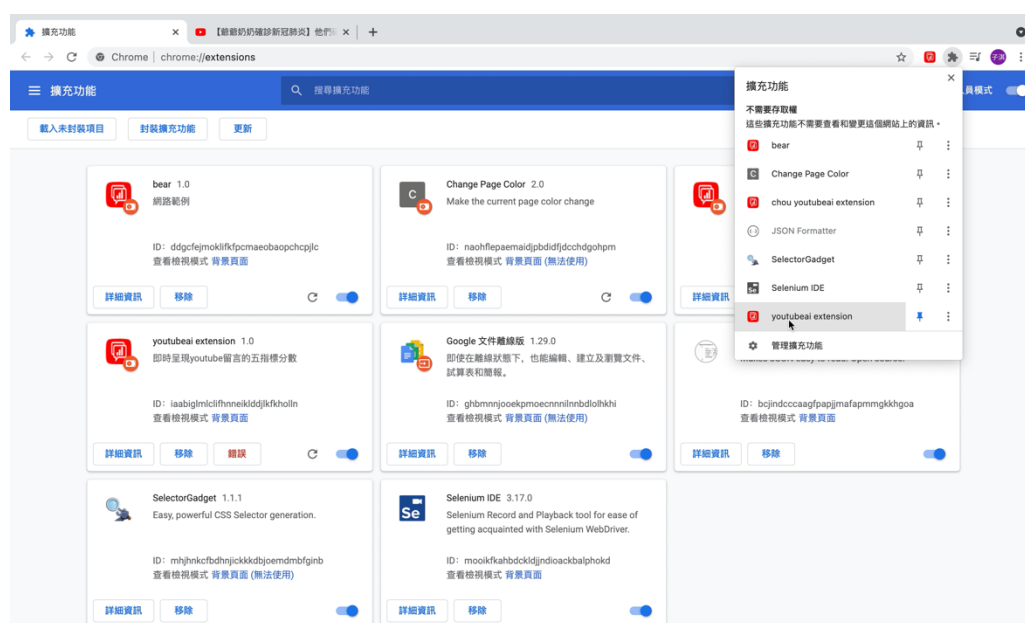


圖 3-2 擴充功能安裝完成後會出現在右上列表

2.使用方式：進入 youtube 影片頁面，將滑鼠點擊右上角 logo 圖示，結果就會呈現在留言上面。

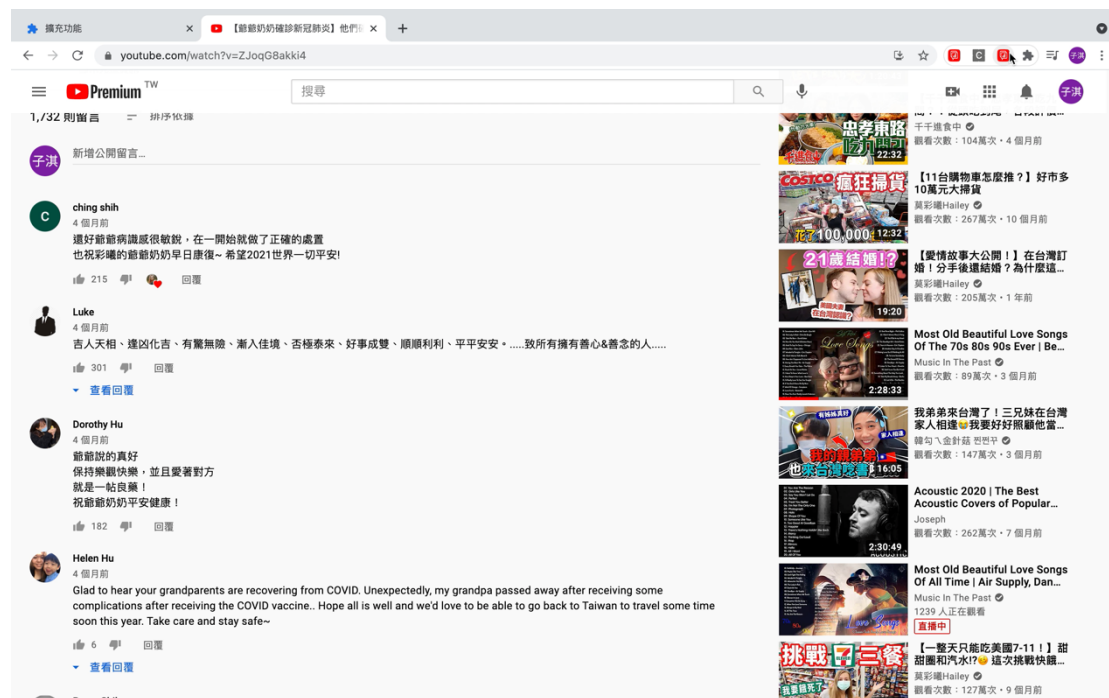


圖 3-3 進入 Youtube 影片頁面，點擊右上角 logo 圖示即可使用

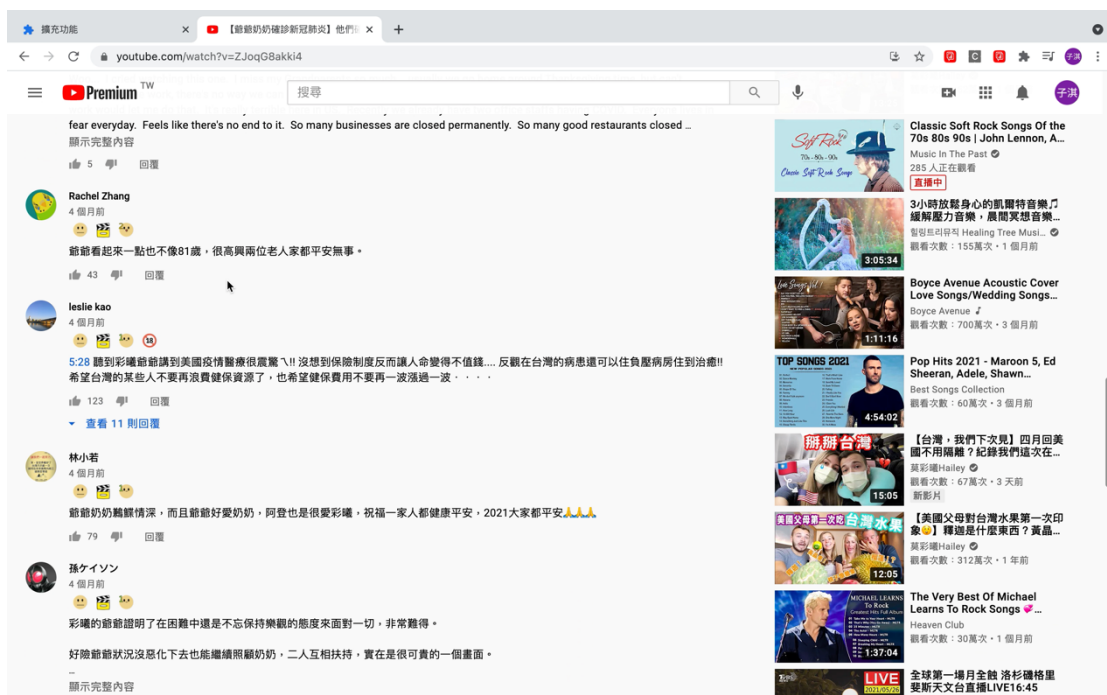


圖 3-4 分析結果就會呈現在留言上面

結論

先前製作的分析網站，必須要使用者提供網址才能在該網站顯現分析情況。本次專題開發出 Chrome 套件，就是要解決不便之處，想讓用戶在使用上更加便利，優化整個操作體驗，所以以 chrome 套件為目標下去開發，讓用戶在 Youtube 頁面即可看到透過視覺化呈現的留言指標。原本一開始有規劃要設計帳號登入的相關功能，但後來因為時間因素將此功能拿掉，是這次專題中比較美中不足的地方。

未來希望可以將原本設計的帳號功能重新完成，讓使用者僅能查看自己影片的留言分析，另外提供即時修正留言指標分數，將模型進行優化，來提高準確度，讓使用者對於分析品質更滿意。同時也以此套件作為一個媒介，來推廣我們這個服務，讓更多人認識。新用戶看完我們套件的圖示後，如果還想要得到更多更專業的分析，就可透過套件導引至網站，來提供更細膩、完整的數據與圖表。

附錄

參考文獻

許愷洋（2019）。跨代收視習慣大不同，長者網路收視率大漲。iSURVEY 東方線上。取自 <https://www.smartm.com.tw/article/36303632cea3>

鄭伊庭（2017）。2018 影音行銷怎麼做？先記住 YouTube、FB、IG「5、1、10」準則！。SmartM。取自 <https://group.dailyview.tw/article/detail/142>

Chrome Extension 開發與實作。iT 邦幫忙。取自 <https://ithelp.ithome.com.tw/articles/10186039>

StatCounter Global Stats Browser Market share Taiwan from Apr 2020～Apr 2021。StatCounter。取自 <https://gs.statcounter.com/browser-market-share/all/taiwan>

StatCounter Global Stats Browser Market share worldwide from Apr 2020～Apr 2021。StatCounter。取自 <https://gs.statcounter.com/browser-market-share>

KD Chang（2017-06-03）Python Web Flask 實戰開發教學 - 簡介與環境建置
取自 <https://blog.techbridge.cc/2017/06/03/python-web-flask101-tutorial-introduction-and-environment-setup/>

網站圖表參考 <https://www.highcharts.com>

網站 loading gif 參考 <https://loading.io>

YouTube 爬蟲參考 <https://github.com/egbertbouman/youtube-comment-downloader>

網站設計 CSS 參考 <https://getbootstrap.com>