〈轉載自清流雙月刊 106 年 11 月號〉

法務部調查局資通安全處　雷喻翔

社群網路與情資蒐集

隨著網路技術普及再加上無線通訊的發展，人手一機走到哪瀏覽到哪的現象隨處可見。當單純地瀏覽網頁已不能夠滿足一般使用者的需求，隨之而來的 是提供使用者互動交流的社群網路應用服務，讓使用者無論身在何處皆可與親 朋好友分享生活周遭的一切。然而當原本立意良善的科技發展被惡意人士濫用 時，所衍生的後果往往需要花費更多工夫加以補救。

近幾年恐怖分子活動頻繁，例如伊斯蘭國（IS）及蓋達組織（AI-Qaeda），其間的串聯與活動亦搭上了社群網路的發展，在其推波助瀾下，政府當局不得不投注心力想方設法遏止這股潮流。身為社群網路龍頭之一的臉書（Facebook），也宣示極力發展 AI（Artificial Intelligence）人工智慧技術以偵測恐怖分子於社群網路上散播之惡意訊息。在此背景之下，本文旨在介紹當前社群網路技術發展及情治機關情蒐方向可能之發展。

社群網路簡介

社群網路興起於 2002 年初期，當時由於 Web 2.0 的發展，

使得網頁應用程式不再侷限於讓使用者單方面從伺服器端接收資訊，使用者開始可以與網頁產生更多的互動，甚至可以藉由網頁與其他使用者產生連結，逐漸地建構了社群網路的雛形。所謂的社群網路其實就是現實生活中的人際關係及虛擬的網路世界兩者的互相投射與聯結。以現在全球最受歡迎的社群網路臉書為例子，使用者可以透過臉書應用程式與朋友互動，傳送彼此間溝通的訊息、推薦喜歡的餐廳、評價某部電影的內容等。藉由這些互動，臉書可以將使用者與擁有共同興趣或喜好的第三人串聯在一起，逐步擴大自己的生活圈。這種自動推薦朋友的方式很容易將一群人連結成一個群組或社團，以往我們熟知的實體社團已被這種虛擬的社團所取代，當要號召同一社團內的成員集體行動時，不再需要為了聯絡 方式而傷透腦筋，只須在網路上一個簡單 的訊息傳遞便可以達到一呼百諾的情勢。2011 年初的茉莉花革命以及 2014 年在臺灣 發生的太陽花學運都是藉由社群網路凝聚了龐大的人潮。

然而原本立意良善的技術經由有心 人士的利用，瞬間變成了犯罪者的溫床。 恐怖分子便相中了這塊執法者公權力尚無 法直接介入的祕境，開始進行了組織性的 宣傳及招募新血，誓言將他們所宣稱的聖 戰傳遞到世界上的每一個角落。透過網路 的

渲染，惡意的假消息及串連活動防不勝 防。既然這是由於社群網路技術發展所致， 當然也應該由社群網路技術本身來加以制衡，這也就是最近幾年社群網路大廠競相 投入最重要的研究議題之一。

社群網路語意分析

語意分析是自然語言處理（Natural Language Processing）的應用之一。自然語言處理簡而言之就是教電腦可以自動識別人類使用的文字，應用的層面包含自動翻譯、文字校對、自動問答系統及語意分析等。語意分析除了讓電腦自動識別文字內容以外，還需要進一步理解文字背後所隱含的意義。例如：「科比·布萊恩全場投 20 中 1，不愧是打鐵界的一哥」這句話看似稱讚，其實是在暗諷投籃命中率過低。

語意分析便是能夠自動察覺使用者實際所要表達的意思。社群網路由於連結全球數以萬計的使用者，每天產出的資料量呈現爆炸性的發展。這些巨量資料裡頭或許隱藏著恐怖分子間溝通的暗語，或是不法分子意圖攻訐政府所散布的假新聞等，這些資訊對於執法機關而言都是相當珍貴且有 用的訊息。

然而如同在砂礫中找尋珍珠般，這項 工作無法單靠人力來完成，必須仰賴電腦 的人工智慧來發掘不易發現的線索，隨之 而來的技術就是社群網路語意分析。一套 完整的社群網路語意分析系統包含下列 4 項步驟：資料蒐集、資料整理、資料分析 及資料呈現。

資料蒐集其實就是網路爬蟲，針對我們鎖定的目標，持續性地透過程式蒐集相關資料。現在熱門的社群網路，像是臉書及推特（Twitter），皆有推出相對應的應用程式介面（Application Programming Interface,API）方便研究人員擷取該網站上的資料，至於沒有提供 API 的社群網站，例如 PTT（批踢踢實業坊），我們可以使用專為處理大量資料而開發的程式語言（Python 或是 R language）提供的套件輕鬆蒐集資料。以往繁複的資料蒐集工作並沒有隨著資料量指數型遞增而越發困難，相對地開發者則是擁有越來越多便捷的開發工具。

資料整理則是在做資料清洗（Data Cleaning）。第一階段蒐集完資料後，並非 所有資料都是有用的，必須汰除不需要以 及冗餘的資料，並且將資料正規化以利下 一階段資料分析。

資料分析是關鍵所在，現在最受歡迎的技術是基於類神經網路（Neural Network）而發展出來的深度 學 習（Deep Learning）

技術。拜AlphaGo屢屢擊敗圍棋高手所賜，深度學習這個詞彙常見諸於新聞媒體上。身為資料探勘(Data Mining)中分類(Classification)技術的一種，深度學習並不限於下圍棋的應用上，而是可以廣泛地運用在任何需要依靠電腦自動判斷並分類的應用中。如前述文章中所提，語意分析是當電腦讀取一段文章後自動判斷出文章的屬性。舉例來說，我們想要做的分析是判斷某一段文章是否與恐怖攻擊有所關聯，所以當電腦讀了一段文章後，程式必須決定該段文章是屬於恐怖攻擊正相關屬性或者是低相關屬性，按照這種模式，我們可以把這個問題歸類成分類的應用，於是乎深度學習便躍於資料分析的舞臺上。

資料呈現主要著重在資料視覺化呈現。當完成上述 3 個步驟之後，最後只差如何將結果具體且顯著地讓第三方知道。傳統報表式的數據雖然可以做到不失真，但是失去了吸引讀者以及達到一目瞭然的效果，於是乎以圖形取代冰冷冷的數字已經成為一種顯學。幸運地，資料視覺化相關的套件有如雨後春筍般一個接著一個推出，從最基本 Excel 所提供的製圖工具到專 為資料科學而生的程式語言（Python 或 R Language），甚至是網頁程式的 D3.js（Data-Driven Document JavaScript package）都可以輕鬆地輸出專業且美觀的視覺資料。

結語

隨著犯罪智慧化，網路已經成為犯罪的重要媒介之一，因此網路情蒐也是接下來重要的犯罪預防手段，若能早期發現可疑的犯罪行為，執法機關才得以籌劃對策，防範於未然。社群網路的蓬勃發展已是不可擋的趨勢，當然它提高了犯罪的隱蔽性，但同時也提供了執法者發掘案源的全新管道,若是能夠善用社群網路情蒐這個工具，原本看似危機的新科技，換個角度，卻是嶄新的轉機。