

經濟史專題導讀__Emotion Analysis for Gig Data

R05323026 經研一 曾韋智

What is the question of the paper?

作者想要做的是利用大數據分析的技術，了解情緒如何在文字中表露，尤其是在社群網站如此發達的情況下，此技術很有發展的空間。並在了解這些帶有情緒的詞後加以應用。

Why should we care about it?

為什麼情緒很重要？為什麼要知道別人的情緒？

情緒在生活當中是不可或缺的。人在相互溝通時總是會帶一點情緒，情緒就像是調味料幫言語加上了額外的意義，在適當的言語加上情緒會有加分的效果，例如老師在叫學生們「安靜」時，帶有生氣的情緒勢必會有加以嚇阻的效果。但同時，文字也是情緒表露的媒介，當我們得以運用數據分析的方式，了解文字與情緒之間的關連，則此些研究發現便可能得以被廣泛運用，無論是商業領域，社會議題的探究，心理健康照護的提升... 等等。

What is your (or the author's) answer?

作者使用了群眾的潛意識的情緒用字，加以處理與排列，以便得知話語中帶有的情緒，和以往的自然語言處理法、資料探勘和機器學習法有所不同。並且再將其運用在兩項分析中：一項是分析 twitter 對蘋果手機的情緒 hashtag，讓時間地點還有評價做結合，得到一筆有價值的資訊讓廠商做進一步的決策。另一項是研究擁有精神疾患(躁鬱症和邊緣性人格症)和其他正常人的情緒特徵詞以及其頻率等，或許能偵測出其他人是否也擁有徵兆，加以追蹤。

How did you(or the author) get there?

作者利用社群網站的免費資源，分析發文者的情緒，像是 hashtag 以及表情符號能得知發文的的情緒。加上由於是大數據的資料，所以可以處理掉一些太短、太複雜或是有 URL 的資料去蕪存菁。再比較非情緒的樣本和帶有情緒的樣本（實驗組和對照組），找出相異和相同處，並以關係圖的樣式建立出兩種圖，再刪除掉重疊的部分，找出關鍵的字詞和常與其連用的字，便可組成情緒特徵詞。再依頻率、多元性等特性加以排序，便可分類成各種情緒類別。作者也提到這方法的準確度高達 80%。