

# 大數據下的情緒分析 - 導讀

R05323023 經研一 吳俊信

## 1. What is the question(of the paper)?

以往對於情緒分析的處理方式有兩種：

(1)自然語言處理法：也就是分析字的詞性、了解字的意義及了解字詞間的關聯。但實際處理不易外，字句字詞換個位置後，意思也不盡相同，而且換個語言就要重新做，不甚方便。

(2)資料探勘/機器學習法：也就是透過搜集大量資料，再以人工去標記，最後透過資料探勘及機器學習分析資料。但該方法也遇到一樣的難題，換個語言就要重新人工標記資料，成本太高，此外，人工標記的資料也不太符合情緒的實際狀況。故，還有什麼方法可以分析情緒呢？

## 2. Why should we care about it?

由於我們對於情緒反應還不夠了解，例如小孩為什麼會哭、或小孩不會字詞、不懂語言，但卻能夠知道其他人的情緒起伏，又或是有時人們生氣的原因，千變萬化。所以如果我們若能透過大量數據來分析情緒，將有助於我們了解人的情緒反應，進而採取最好的回應策略。

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

研究者發現，情緒是從許多小細節累積而成，但過去的分析方式往往忽略這些小細節。所以研究者想到從網路上找尋合法的資料（例如維基百科、臉書、推特等等），進而從人類的日常生活紀錄中，搜集資料。例如發文時常常會出現的 hashtag 及表情符號，透過這些資料，來判讀情緒。

## 4. How did you(or the author)get there?

研究者透過實驗組-潛意識情緒資料(例如個人臉書發文)，及對照組-非情緒資料(例如CNN發文)，分析字詞出現的頻率、分析字詞互相伴隨出現的頻率及比較字詞間的重要關係程度等等，進而找出其相同、相異處，接著建立兩種情緒圖，將兩圖相同的地方剔除，接著分析哪些字是所謂的中心點、哪些字常常被一起使用，留下圖形中的重要字詞，再將字詞做情緒特徵分類，最後再將特徵值送進分類法學習，即可判讀情緒資料。