

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

Big data 的情緒分析。有許多個人的原因，像是在 Intelligent Data Engineering and Application Lab 的一位國際博士生表示：有時候我們比較不容易明白女性朋友說的話，可能比較模糊或是模稜兩可。而尚未學會說話的孩子，只會以哭喊來表達需求，也很需要我們傾聽他們的心靈。透過情緒分析 (Emotion Analysis) 能讓我們更瞭解表達者的心思意念。

### Why should we care about it?

過去使用自然語言處理 (natural language processing) 與資料探勘的方式，自然語言處理需要字典以及語意關係、字的位置也影響句子的意義、且不同語言需要不同的資料。資料探勘則是針對不同語言，需要培養不同的資料。過去的方法忽略瑣碎的字詞，像是停止詞 (stop words) : often, above, again，限定詞與代名詞也常被忽略。錯字並沒有被記錄在字典中，但有些錯字是故意且能表達情緒的，比如 soooooo happy。

情緒分析像是師法尚不懂文法與詞彙的一歲小孩偵測媽媽情緒的方法，目標是去除文法、字典、錯誤與語言的限制。

經費不足可透過許多網上的免費資源與群眾外包 (crowdsourcing)，利用網路將工作分配給大眾。

### What is the author's answer?

案例一：由情緒分析得到消費者對商品的喜好

案例二：由個人情緒狀態分析找出誰患有躁鬱症與邊緣人格障礙  
也許有機會用於選舉分析

### How did the author get there?

案例一：利用潛意識的情緒特徵進行文字內容分析

主題標籤 (hashtags) 以及表情符號都很能表達情緒

1. 預備步驟：移除棘手的部分，像是太短而不重要的資訊、有太多主題標籤 (可能資訊過量)、是轉貼的訊息、有網址 (不易取得此網站上的資料)。
2. 找出實驗組及對照組共同與不同的地方。
3. 畫一個包含情緒 (emotion) 的圖與一個沒有情緒 (non-emotion) 的圖，比較兩個圖共同的地方，然後只留下包含情緒圖中與無情緒圖中不同的部分。接著，運用社會網路的方法分析文字的集中度 (centrality) 與群聚度 (cluster degrees)。
4. 將情緒樣式依據頻率、排他性與多樣性排序，得到每一種情緒的排序清單。

案例二：分析個人的情緒狀態找出誰患有躁鬱症與邊緣人格障礙

1. 經由社交網站（**twitter**、**fb...**）蒐集病人與正常人的資料
2. 找出實驗組及對照組共同與不同的地方。
3. 利用擷取的特徵進行分類：神經網絡、原始貝氏（**Naïve Bayes**）以及（**Bayesian Belief Networks**）貝氏信念網絡、支援向量機（**Support Vector Machines**）、隨機森林（**random forest**）演算法