

Emotion Analysis for Big Data

經濟史專題討論

許芳瑀 2017.3.16

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

社群網站充斥各種情緒，此研究欲探討如何從大數據中有效分析潛意識裡的各種單詞以理解發文者的情緒；並瞭解這些情緒特徵詞後可如何運用。

2. Why should we care about it?

有時很難從他人的言語中讀懂情緒，不管是親密的朋友、家人或是牙牙學語的嬰孩亦是如此，因此希望藉由情緒特徵詞來了解個人甚至群體的想法，讓調查者直接運用大量的詞意直接了解發文者的情緒及想法，省去各式訪談或是調查所耗費的時間成本。

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

相較於過往的自然語言處理與機器學習，此研究藉由淺意識情緒字眼整理與再排列來分析文字中隱含的情緒。而之後其運用有：

一、在商品的喜好分析中，藉由 iPhone 的標籤(hashtag)了解發文者的所在地與對此項商品的評價，使廠商有更精準的策略方針。

二、藉由收集大量躁鬱症(bipolar disorder)或是邊緣性人格障礙(Borderline Personality Disorder, BPD)病患與正常人的樣本後，分析其情緒特徵，來預測其他發文者是否有潛在的病徵。另外探討是否能對於選舉有更進一步的掌握。

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

現今人們喜歡在社群網站上抒發自我情緒，本研究藉由這些公開免費且易取得的群眾資料進行分析，其中標籤與表情符號能有效標註發文者的情緒，因此作為被標記的關鍵詞，並將資料分成具有情緒資料的實驗組與非情緒資料的對照組。然而在實際文章中有許多混雜的資訊，故在大數據有足夠大樣本之下，選擇去掉一些麻煩的資料，例如太短無法判斷情緒、擁有太多標籤、轉寄、有額外超連結.....等。

利用圖表繪製出實驗組與對照組的相異處，找出其關係程度，再移除掉重疊的部分，即能找出只有情緒資料所能獲取的重要詞彙，利用圖形資料庫繪製出主要的情緒特徵，依照其頻率、特性、多元化，將其分門別類歸納成各種情緒。

最後，運用此方法探測情緒分類，其準確度皆有達到八成，因此認為此方法可行。