

(a)

資料集：

Sarcastic Comment – REDDIT

內容：

這個 **dataset** 包含一百三十多萬筆來自國外知名論壇 **Reddit** 上的諷刺性留言。判斷依據為留言是否帶有 **\s (sarcasm)** 的 **tag**，**Reddit** 上的用戶們使用這個 **tag** 去代表自己的留言是屬於開玩笑、帶有反諷意味的，不需要大家認真看待。相比於自行判斷、留言自帶的 **tag** 是個相當可靠的依據。

此 **dataset** 亦包含非諷刺性的留言，每筆 **data** 除留言內容外也包含：留言所對應的文章主題、被其他用戶噓或推的數量、所討論的領域等等。

問題定義：

我們是否能利用上述資料訓練，建立出一個 **model** 去判斷一段留言是認真回覆還是反諷意味。

潛在問題：

要辨別認真和反諷存在許多模糊地帶，但要表達的結果往往是兩個極端、一旦判斷錯就會是完全相反的結果。例如：“看看你做的好事！”，在中文中看起來應該要是稱讚意味，但實際應用上的意思卻恰好相反。

分析與預測難度：

要判斷出一則留言是認真還是反諷，需要透過文字的片面資訊去揣測發文者的原意，即使是對情境判斷及文意理解能力強的人腦都可能常搞錯，對電腦來說勢必也是一大挑戰。可能的分析依據會像是：刻意全大寫、斜體、加粗、拉長或帶有引號的字詞，例如：“Yeahhh, I'm sure THAT is the right answer”。

價值：

在這資訊發達的時代，網路論壇上大家所表達的意見已是不可忽視的聲量，各領域可能都需要透過網路上用戶們的回饋去調整或改進自身的做法，若能精確的判斷諷刺性回覆背後的本意勢必是一大優勢。

舉最近政治選舉為例，許多論壇上的網友可能會透過一些玩笑性的稱讚獲推舉去譏諷特定候選人的政見或作為，並引起大家群起效尤，亦即所謂的反串粉絲。若能精確判斷大家的本意，將有利於媒體做出更精確的民調或

網路聲量調查，以及有益於候選人陣營去評估該繼續進行的策略與調整。

(b)

我自己本身是電機系的學生，在電機領域當中目前很火紅的領域之一是與生醫結合，研發出智慧型晶片或穿戴式裝置等等。如果能收集大量有關生理訊號的資料並建立資料集，應該能在居家照護方面等等有相當多的應用。

像是可能藉由收集嬰兒哭聲與記錄是何種因素使其滿足並停止哭泣，例如：尿布濕、肚子餓、想討抱抱等等。若能發現各種不同狀況嬰兒的哭聲形態特徵不同的話，我們或許能建立 **model** 去分析判斷嬰兒每次哭泣時所對應的背後原因，並依此滿足其需求，對症下藥。

或者我們可能藉由穿戴式裝置或影像處理的方式，去記錄家中老人行動時肢體活動的資料。若發現某次出現異常的活動行為模式，很可能就是發生跌倒或其他突發狀況，若能以此資訊快速發出預警或通知親友，就可以將傷害降至最低。