

Analysis and Application of Natural Language Processing to Mental Illness

自然語言處理對心理疾病之分析與應用

組別：B133 指導教授：陳宜欣教授 組員姓名：馬禹平

Abstract

本研究旨在利用自然語言處理技術判別社群媒體上的文章是否具有憂鬱症或自殺自傷傾向。根據聯合國世界衛生組織(WHO)統計，全球約有 2.64 億人罹患憂鬱症。憂鬱症不僅造成了患者情緒與經濟上的負擔，更影響了整體社會生產力。在今日社會中，社群媒體提供了人們抒發心情的管道，甚至人們在社群媒體上的發言比日常生活更接近其內心想法。基於上述情形，本研究希望利用社群媒體中與憂鬱情緒或行為等表現相關的貼文作為研究資料來源，利用深度學習將文章中的句子分類，進而判斷文章發文者是否具有憂鬱症、自殺自傷傾向，建立憂鬱症危機文字自動判別系統。

本研究利用雙向長短期記憶模型(Bi-LSTM)加上 Attention 機制作為主要模型架構，進行文句的分類。同時，本研究亦嘗試了 Doc2Vec 模型和其他分類器如支援向量機(SVM)作為分類模型，並比較不同模型差異。文章分類模型將基於分類後的文句資料進行訓練。本研究透過不同類別中所含有的句子數量進行統計，利用不同的分類器如 Adaboost、決策樹(Decision tree)、羅吉斯迴歸(logistic regression)模型對文章進行分類，判斷文章是否具有憂鬱症或自殺自傷危機，並比較不同模型間的關係並找出可有效分類文章是否具有危機的模型。

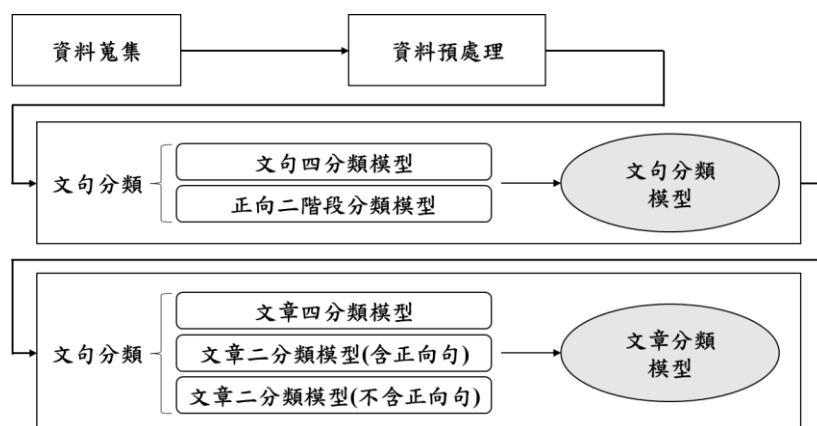
根據研究結果我們發現，影響文章是否具有危機的主要因素為文章中所含有的負向句的多寡。在文句分類模型中，Bi-LSTM+Attention 的三分類模型有最佳效果；而在文章分類則是使用羅吉斯迴歸分類模型可達約八成的準確度。同時我也發現，針對情緒相關的文字若使用傳統的斷詞方法，將無法有效判斷文句所表達的情緒。若要對此改善可能需要建立情緒專用的語料庫協助進行文字清理。

綜上所述，本研究最終將結合文句分類模型和文章分類模型，用以分析社群媒體中的貼文是否具有憂鬱症或自殺自傷危機。同時本研究也提出了未來改善方式，期望建立情緒語料庫來協助對情緒文本的辨別與分析。本研究希望最終成品可實際被使用，以偵測潛在的憂鬱症患者並提供協助。

Implementation

I. 研究流程

圖(一)為本研究的研究流程，本研究將參考文獻中的方法並加以調整，以符合本研究的目的，訓練出可分類網路危機文章的模型。本研究將主要分為三階段：第一階段為資料蒐集與預處理；第二階段為建立文句分類模型，並比較模型間的差異與優缺點；第三階段為建立文章分類模型，利用第二階段所的結果將文章進行危機程度的分類。



圖(一) 研究流程

II. 模型介紹

(I) 文句分類模型

本研究將文句分為以下四類，分別為正向、負向、生理反應和自殺行為。由於文句資料具有數量不平衡的現象，其中正向句與自殺行為句的數量遠少於其他二類，因此，本研究將嘗試兩種分類方法。第一種為直接將文句按照四類進行分類。第二種則是進行複合式分類，透過結合在不同分類狀況下具有良好表現的模型對文句進行分類。

1. 四分類模型

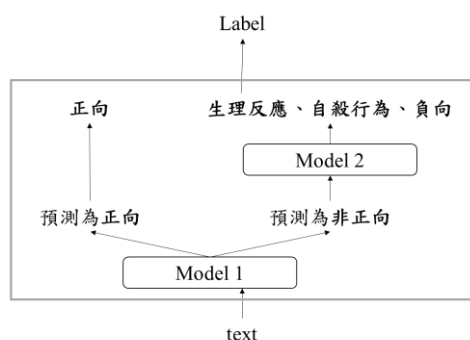
本研究利用了不同的分類模型進行測試，其中在四分類模型部分表現較佳的模型為利用 TF-IDF 搭配羅吉斯迴歸模型。然而由於除了正向以外的三類別皆具有非正向的用字、文意，因此模型在學習時容易被混淆。對此，我提出了正向二階段分類模型，以期改善此狀況。

2. 正向二階段分類模型

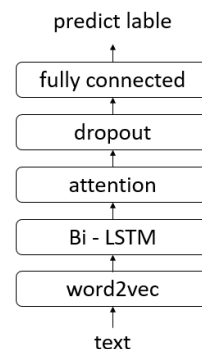
圖(二)為正向二階段分類模型，此模型中 Model 1 將會先將句子分為「正向」和「非正向」兩類，再進一步使用 Model 2 將「非正向」的句子分為負向、生理反應和自殺行為三類。在研究過程中發現，Model 1 的分類結果並不如預期好，若和 Model 2 進行串接只會使整體分類效果更差。在 Model 2 中，利用 Bi-LSTM+Attention 模型(圖三)可達到最好的分類效果(準確度約七成)。

3. 其他討論

在分析文句分類結果時我發現，在文句預處理中的移除停用詞與結巴斷詞時，會將一些影響文句意思的副詞移除，造成結果與原本句意有落差或相反。此外，文句中的連接詞和語氣助詞也會影響文句所展現出的情緒，但卻會被移除。因此，在處理情緒相關文字時，除了不移除停用詞外，也應該考慮建立適用於情緒文字的語料庫來進行斷詞或移除停用詞。



圖(二) 正向二階段分類模型圖



(三) Bi-LSTM + Attention Model

(II) 文章分類模型

標注資料將文章分為高度危機(A)、中度危機(B)、低度危機(C)和無危機(D)四類。考慮到目前測資可能過少而導致分類效果不佳，我嘗試將高度危機和中度危機合併為「危機」、低度危機和無危機合併為「無危機」。

基於文句分類模型的結果，文章分類模型將以「正向句數」、「生理反應句數」、「自殺行為句數」、「負向句數」和「情緒句總句數(以上四項句數加總)」五項特徵代表文章的輸入進行分類。文章分類模型將利用不同的判別式模型如 SVM 模型、決策樹模型、隨機森林，Adaboost 模型等，以及生成式模型如單純貝式分類器進行分類，並比較其之間差異，已找到最適合此資料的分類模型。

1. 四分類模型

此部分如實驗前所預期，即使是表現最好的模型也只有約六成的準確度。

2. 二分類模型

在此模型中，各種分類器皆達到良好的分類效果(約八成)。然而文章分類模型是基於文句分類的結果再進行分類，因此在與文句四分類模型連結後期表現效果可能大打折扣。考慮到本研究最終目的為進行文章危機性的判別，因此決定將文章分類模型中「正向」這項特徵移除，「情緒句總句數」改為三類非正向句數量加總。在進行訓練與測試後，發現其表現與舊有的模型相差不大，亦具有約八成的準確度。因此最終決定捨棄正向句特徵，單純使用非正向三類進行分類。

3. 其他討論

在觀察原始資料和分類錯誤資料時可發現，在「無危機」類別中仍有文章含有較多的非正向句；在「危機」類別中亦有含有少量負向、自殺行為句的文章。

對此我認為文句分類和文章分類結果不一定直接相關，可能為分類後的文句仍具有情緒程度上的差異造成。

III. 結論

本研究最終結果顯示利用 Bi-LSTM+Attention 模型建立的非正向文句三分類模型為文句分類模型中效果最佳的分類模型；文章分類的部分則是配合文句分類模型，選擇基於非正向句的文章二分類模型。

另外，本研究亦發現對於與情緒相關的資料，並不適合使用傳統的文字預處理方式。對此除了採用關注前後文語意的 Bi-LSTM 模型和 Doc2Vec 模型外，也因考慮建立適用於情緒文字的語料庫，以提供往後需處理此類型資料時有更完善的資料預處理方式。

心得

會想進行這個專題研究的原因是因為我希望能將我在電資領域所學到的知識應用於協助解決社會問題，而憂鬱症又剛好是現今對人類健康影響很大的心理疾病之一。在研究過程中，我學到了很多機器學習的知識，而針對自然語言處理方面也透過閱讀各式文獻與參考書籍學習到了很多這方面的應用。很感謝宜欣老師對於我研究方向的指引，也在我遇到困難時提供解決問題的思路，從旁引導我在研究上有更全面的探索。

另一方面，由於此研究涉及了心理諮商的知識，雖然資料的標註並不是由我負責，但在研究過程中亦需要對資料有一定的認識，才能去選擇更適合的方式。在研究中遇到最大的困難就是資料標註可靠度的問題，例如一開始文句分類表現很差，在和心諮系的老師討論後發現原本的標註方式不適合用於機器學習，因此重新進行標註，也浪費了一點時間。這次的經驗也讓我了解到在開始訓練模型前須先確認好資料來源的品質，去了解的性質。最後，要謝謝宜欣老師和其他在研究上提供協助的老師們這一路上給我的幫助，耐心地指導我完成這個專題研究，使我在研究的過程中獲益良多。