標題：善用模型特徵－從文字預測到情感分析

關鍵字：非監督式學習、文字預測、文本情感分析、mLSTM

監督式學習(supervised learning)是目前推動深度學習的主要動力，然而蒐集資料易，標記資料卻相當費神，僅有少數足夠重要的任務才值得我們大規模投資，例如：影像分類、語音辨識與機器翻譯等等。於是，科學家一直致力於非監督式學習的相關研究，期待能在訓練過程中，大幅減少需要人工標記的數據，像是利用已開發模型抽取資料庫的特徵，進行另外一段非監督式的任務。

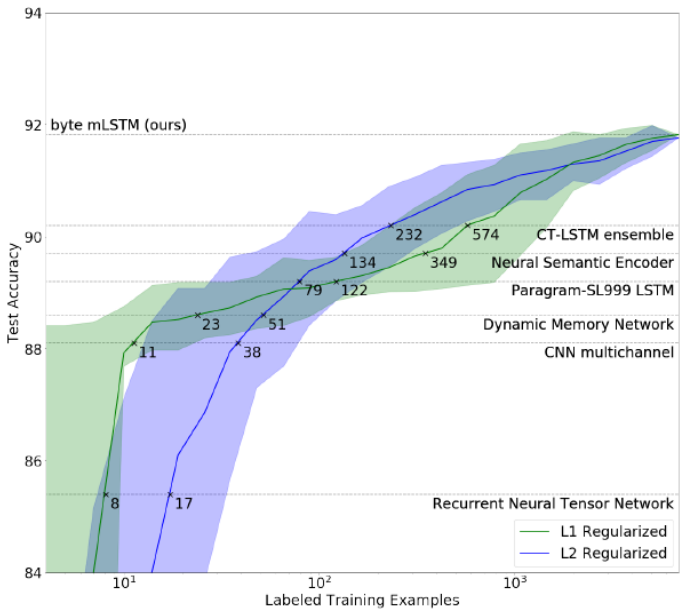
文本情感分析

一個乘法長短期記憶模型(mLSTM)，在歷經一個月的訓練，看過大約八千兩百份 Amazon 顧客評論之後，能夠實現單詞的文字預測。而其中包含的 4096 個單位，等同於一個字串的特徵向量，當我們深入研究，找出這些單位最有效的線性組合，便有望將文字預測模型轉為情感分類器，以字元為單位逐步分析文句的論述態度。

METHOD	MR	CR	SUBJ	MPQA
NBSVM [49]	79.4	81.8	93.2	86.3
SKIPTHOUGHT [23]	77.3	81.8	92.6	87.9
SKIPTHOUGHT(LN)	79.5	83.1	93.7	89.3
SDAE [12]	74.6	78.0	90.8	86.9
CNN [21]	81.5	85.0	93.4	89.6
ADASENT [56]	83.1	86.3	**95.5**	**93.3**
BYTE MLSTM	**86.9**	**91.4**	94.6	88.5

(圖一、各模型對簡易資料庫的分類表現。由於文字在處理過程中需以 UTF-8 編碼表示，故我們的模型又被稱為 byte mLSTM)

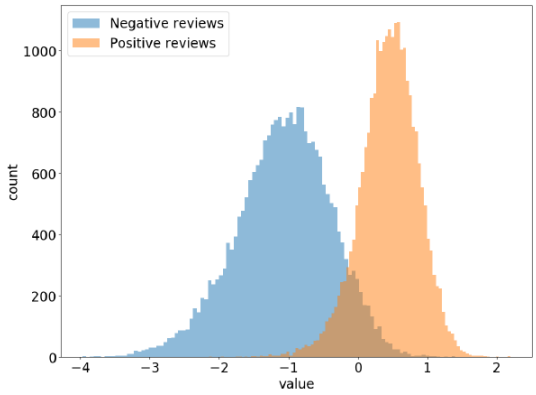
由圖一所見，我們的模型(即 byte mLSTM)在四個簡易的標準資料庫中，有著不一樣的表現。由於 MR 和 CR 是由電影與產品的評論網站中蒐集數據，涉及領域與我們的訓練資料庫高度重疊，因此非監督式的文字模型從中學習到豐富的文本特徵，使我們的情感分析模型獲得壓倒性的勝利。然而，在判斷主觀性/客觀性的 SUBJ 以及分析意見傾向的 MPQA 之中，byte mLSTM 則不敵監督式模型，甚至其他知名的非監督式模型(如 Skip-thought)。



(圖二、byte mLSTM 與各模型在 SST 上表現的比較，其中 byte mLSTM 分別使用 L1 正規化與 L2 正規化)

接著，我們將模型推向 SST(Standard Sentiment Treebank)資料庫，挑戰更複雜的文本結構。或許是因為 SST 與 MR 的取材類似，byte mLSTM 以準確率 91.8%刷新最高紀錄；並由圖二可知，byte LSTM 運用資料的效率極高，光靠幾百筆的標記資料，便能凌駕於他人之上，而值得注意的是，這些訓練樣本根本不及 SST 的十分之一，甚至與其他模型相差 30 至 100 倍。

情感神經元



(圖三、情感神經元面對正面評價與負片評價時的數值統計)

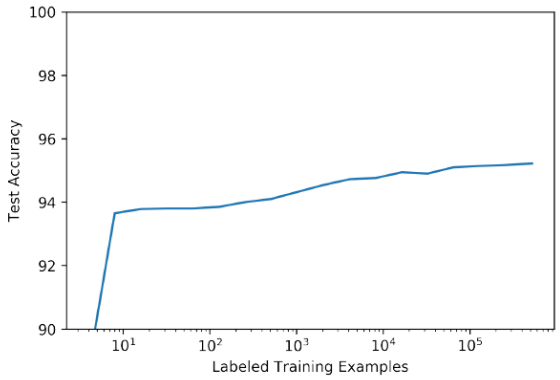
(圖四、byte mLSTM 是逐字進行情感分析，綠色代表正面，而紅色代表負面)

當我們更進一步討論 byte mLSTM 學到的諸多特徵時，發現模型中有個「情感神經元」相當直接地呼應著預測結果。我們利用 IMDB 的影片評論作為測試，在面對大部分的正面評論時，情感神經元會帶著正值，反之則為負值(如圖三與圖四)。而單用情感神經元進行情感分類時，誤差值為 7.70%，表現僅比 byte LSTM 差上 0.58%，代表情感神經元幾乎主宰了模型的預測行為，其中的主要訊息皆由純量承擔，整個過程簡潔卻精幹。

現在我們將焦點轉回建立模型的初衷，既然原先 mLSTM 被訓練為生成模型，那

麼我們應該能夠控制情感神經元的數值，生成符合設定的文章吧！根據圖五的結果，情感神經元的正值確實能夠招致正面詞句，負值也能產生負面詞句。然而在負面樣本中，難免參雜著一些正面語氣，或許這是因為訓練資料庫中，正面評價約為負面評價的五倍，才導致模型在判斷情感或生成文字時有類似的不公之舉產生。

模型表現與延伸



(圖五、byte mLSTM 在 Yelp 資料庫中有明確的能力上限)

為了更遠的推廣，我們勢必擴大這些非監督式特徵的規模，但由圖五所示，單純增加訓練樣本數可能只是徒勞，造成這樣能力上限的原因有三：

首先，byte mLSTM 僅在 Amazon 顧客評語上訓練，一旦面對像是 Yelp 或其他領域的資料庫，需要考量更多專業因素時，表現固然相當有限。

此外，byte mLSTM 畢竟為字元等級，判斷文章時的儲存配置難免稍嫌吃力，往往只著重最後的幾段詞句，而非整篇文章的語氣，壓縮了誤差的進步空間。或許階層式模型(hierarchical model)能是我們將前進的下個目標，改善時間步驟過長所衍生的問題。

最後，簡單的模型結構並非完美，面對訓練資料的增加，只在靜態特徵上線性組合的表現終究會飽和，反倒是往往有明確目標任務的複雜模型，雖然需要更多的標記資料，卻有望飛速進步，甚至超越我們的模型。

但不管如何，byte mLSTM 確實在特定領域的詞句情感分類時繳出革命性的成績，同時也點亮了善用已開發模型的希望，尤其是這類的一步預測模型(next-step prediction model)。或許在一個能夠預測畫格的巨型神經網絡中，也有值得挖掘的非監督特徵，為物體偵測、場景辨識與行為分類提供更好的發展。

參考資料：

1. Alec Radford, Ilya Sutskever, Rafał Józefowicz, Jack Clark, Greg Brockman, "Unsupervised Emotion Neuron", Open AI, 6 APRIL 2017

2. Alec Radford, Ilya Sutskever, Rafał Józefowicz, "Learning to Generate Reviews and Discovering Sentiment", 6 APRIL 2017