自適應中文維度型情感詞典之建立

An Adaptive Method for Building a Chinese Dimensional Sentiment Lexicon

林應龍　Ying-Lung Lin
元智大學資訊管理學系
Department of Information Management, Yuan Ze University
fxm900206216@gmail.com

禹良治　Liang-Chih Yu
元智大學資訊管理學系
Department of Information Management, Yuan Ze University
lcyu@staurn.yzu.edu.tw

摘要

在文本的情感分析(Sentiment Analysis)的任務中，基於詞典的方法因具有高可解釋性且容易使用，中文維度型情感詞典(Chinese Valence-Arousal Words, CVAW)已是重要的基礎工具，本研究的主要目的則是發展一種自適應方法(Adaptive Method)擴充該情感詞典，使其可擴充並適應到不同領域，故本研究利用深度學習的嵌入(Embedding)技術，從健保領域專家標記結果取得新詞的維度型情感(Dimensional Sentiment)，擴充中文維度型情感詞典為自適應中文維度型情感詞典。為驗證該方法之有效性，我們以中文維度型情感詞典作為基線(Baseline)，並加入支援向量機(Support Vector Machine, SVM)及極限梯度提升(Extreme Gradient Boosting, XGBoost)等熱門演算法進行比較，實驗結果顯示，自適應中文維度型情感詞典在交叉驗證實驗中之均方誤差(Mean Square error, MSE)為 0.95、皮爾森相關係數(Pearson's Correlation coefficient)為 0.71，效能略優於基線及其他機器學習演算法。未來亦將結合標記推薦系統，完善具方向性的標記及學習循環，使自適應中文維度型情感詞典能更有效的持續發展。

關鍵詞：自適應，維度型情感分析，情感詞典，情感嵌入，健保政策

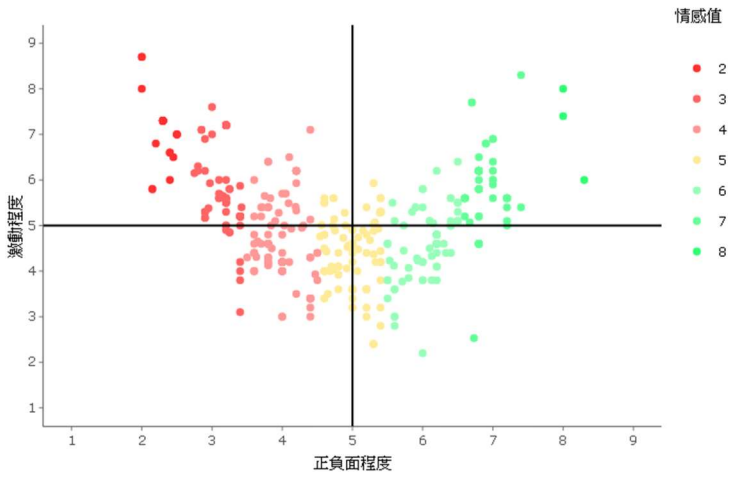
一、實驗目的

因中文維度型情感詞典建立在比較通用的領域中，可能缺少某些公共事務領域的情感詞，導致情感分析的誤差，故為擴充該詞典至不同領域，我們提出一種自適應方法，以專家標記作為基礎事實，透過深度學習的嵌入能力學習新詞的維度型情感，使不同領域可有效持續發展維度型情感詞典，強化情感分析的效能與正確性。

二、文獻探討

在文本的情感分析任務中，大致可分為基於詞典(Lexicon-based)[1]及基於學習(Learning-based)[2、3、4]等 2 種方法，基於詞典的方法因具有高可解釋性且容易使用[5]，亦可融入基於學習的方法[6]，並結合注意機制(Attention mechanism)強化模型的效能[7、8]，因此情感詞典在情感分析領域有著十分重要的地位，已是不可或缺的基礎工具。

故我們在進行中文的情感分析時，中文維度型情感詞典[9]將是有用的基礎工具，而維度型情感如圖一所示，係透過兩個維度之情感，輔助同時判斷情感之正負面及其強度， X 軸為正、負面程度(Valence，1-9 分)、Y 軸為激動程度(Arousal，1-9 分)，X 軸愈高分愈正面，愈低分愈負面，Y 軸愈高分愈激動，愈低分愈平靜。應用該詞典於情感分析時，相較常見的單維度情感正、負面模型，可進一步區分輿情是否激動，並利用兩維度呈正相關的特性，輔助判斷情感分析的正確性，即情感值趨正、負面極值時，通常激動值將較高，反之則較低。



圖一、維度型情感模型

中文維度型情感詞典是從較通用的領域所建立的，在特定領域中則可能因缺乏領域情感詞導致情感分析誤差，雖人工擴充詞典[10]可有效解決此問題，但標記領域情感詞成本較高，因此本研究發展一種可使維度型情感詞典自動擴充及適應到不同領域的方法。

在演算法部份，我們加入支援向量機[11]、極限梯度提升 [12]等常見的機器學習(Machine Learning)演算法進行比較，並使用了深度學習方法[13]進行情感嵌入。支援向量機係為透過核函數(Kernel Function)嘗試將資料從低維度映射到更高維度的空間，使超平面(Hyper Plane)可在映射後的空間最小化誤差，在本研究使用徑向基函數(radial basis function, RBF)作為核函數；極限梯度提升則是決策樹(Decision Tree)集成(Ensemble)及提升(Boosting)的一種方法，透過生成弱學習器(Weak Learner)使決策樹學習資料的某個部分，並透過對誤差的監督決定新的弱學習器是否生成，當弱學習器達指定數量後再透過投票(Voting)機制集成，此演算法對特徵及學習樣本同時進行自動化調整，可使學習更為有效穩定。

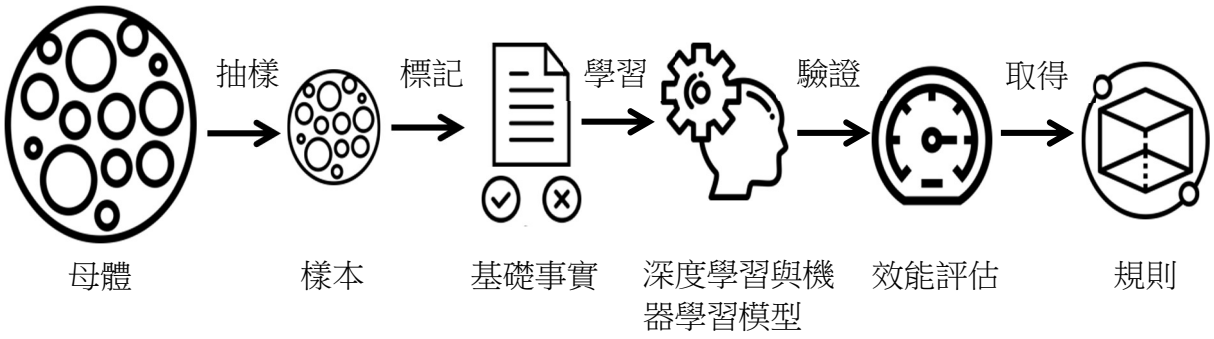
深度學習為一種神經網路架構，其目的為透過不同的架構設計及微調(Fine-tune)[14、15、16、17、18]，使端到端(End-to-End)的倒傳遞(Back-propagation)過程中，自動調整神經元的參數達成最小化誤差，而在架構上則大致可分為編碼器(Encoder)及解碼器(Decoder)等 2 個部分，編碼器部分負責從原始資料中萃取特徵，解碼器則負責從萃取完成的特徵解碼為目標值。因深度學習架構具有編碼器，其透過映射(Mapping)可保留表徵(Representation)，因此擁有優異的表徵學習能力[19、20、21]。

如詞嵌入(Word Embedding)[22、23、24]即為自然語言處理(Natural Language Processing, NLP)領域中應用深度學習取得表徵的重要方法，其透過預測前、後詞的任務調整投影層(Projection Layer)，在訓練完成後該層取出即為詞向量(Word Vector)，若某些詞在訓練文本中的前、後詞相似，則詞向量的相似度將較高，因此保留了前、後詞資訊，相較 One-hot 編碼有更多的資訊量，且因映射到固定長度的向量空間，相較 One-hot 編碼可減少運算量。

而本研究中將藉由深度學習的嵌入能力，應用在自適應中文維度型情感詞典，透過自動編碼從訓練資料中嵌入新詞的情感值，使中文維度型情感詞典可自動擴充詞典並適應到不同應用領域，強化應用範圍及效能。而情感詞典的自適應，已有研究透過基因演算法[25]或深度學習嵌入[26、27、28]，而本研究使用自動編碼[29、30、31]實現情感自適應。

三、實驗設計

為實作自適應中文維度型情感詞典，其流程如圖二所示，可分為建立基礎事實(Ground Truth)、學習規則、驗證規則及應用規則等 4 個階段，第 1 階段將使用抽樣技術取得具代表性之樣本，並進行人工標記，第 2 階段使用機器學習及深度學習演算法學習情感正負面程度與激動程度之規則，第 3 階段則以評估指標驗證規則的效能，最後的第 4 階段則是取得新詞維度型情感並應用在情感分析，透過這 4 個階段的流程，實現可持續自適應，改進情感分析效能與正確性。



圖二、自適應中文維度型情感詞典實驗流程

在資料部分我們透過網路問卷方式取得真實健保政策輿情共 6,920 則輿情，並經過縮減樣本為 1,200 則後，由國立陽明大學醫務管理研究所熟悉健保政策領域的 2 名研究人員標記資料，將以此作為基礎事實，評估中文維度型情感詞典的效能並作為基線。接著我們評估了深度學習與機器學習模型，並從深度學習的嵌入隱藏層(Latent layer)取得新詞，新詞即中文維度型情感詞典中尚未被收錄的詞，透過增加新詞將中文維度型情感詞典擴充為自適應中文維度型情感詞典。

在標記前我們考量人工標記成本高昂，為了使標記工作更有效率，採用了 4 個步驟的樣本縮減方法，第 1 個步驟為去除句子長度低於或高於 2 倍標準差的句子，使句子的長度適中，較適合用於後續標記及學習情感。第 2 個步驟為去除相似句，此步驟係為減少重複標記，以 One-hot 編碼為詞向量後計算歐幾里得距離(Euclidean Distance)，並將小於兩倍標準差的結果視為相似樣本，只保留其中 1 個。第 3 個步驟為符合情感分布，避免隨機抽樣有過度集中在負面或正面情感的狀況，我們先使用中文維度型情感詞典預先計算每一個句子的情感正負程度及激動程度，再分別計算各組別的平均值及標準差等統計值，每組抽樣 100 筆，並以抽樣後的統計值與原統計值計算誤差，作為目標函數，經 1000 次隨機抽樣後取目標函數之最佳結果，此方法確保抽樣後情感正負程度與激動程

度的統計值與原統計值較相近。第 4 個步驟將每組樣本的句子隨機排序後合併，如表一所示，6 個組別在 106 年有 600 筆資料、107 年有 600 筆資料，共 1,200 筆資料。

在經過熟悉健保政策領域的 2 名研究人員標記完成後(結果如表二)，我們使用平均值、標準差及皮爾森相關係數觀察標記結果。皮爾森相關係數可用以觀察兩組標記的相關程度，該係數之區間為 0 至 1，越接近 1 代表越相關。其中情感正負程度的相關程度較高，標記較一致；而激動程度的標記則相關程度較低，從原始標記資料可看出 2 名研究人員的標記較容易出現相反結果，且標記均集中在 4.5 到 5.5 區間，故標準差較低。

表一、健保政策問卷樣本縮減統計

問卷來源	年度	蒐集數	去除句子長度離群值	去除相似句	抽樣
中醫	106	645	611	556	100
	107	547	534	490	100
牙醫	106	698	660	626	100
	107	572	547	511	100
全民健保	106	673	645	610	100
	107	624	591	523	100
西醫基層	106	628	593	562	100
	107	269	256	237	100
門診透析	106	532	502	497	100
	107	527	500	479	100
醫院	106	666	640	614	100
	107	539	515	473	100
總計		6,920	6,594	6,178	1,200

表二、健保政策問卷樣本標記統計

標記種類	樣本數	平均值	標準差	相關係數
Valence	1,200	5.3613	1.4243	0.8672
Arousal	1,200	5.3327	1.0248	0.3313

在斷詞部分我們使用結巴(Jieba)斷詞，並自建含 8 萬多個詞的自定義詞典，其權重以長詞優先，並加入領域相關的特定用語，以提升情感分析的正確性，減少因無法正確斷詞而造成的情感值誤差。在效能的實驗設計部分，我們使用預訓練的詞嵌入進行編碼後，評估極限梯度提升、支援向量機及深度學習等 3 種演算法，亦評估新詞及自適應中文維度型情感詞典等 2 種字典法，並以中文維度型情感詞典作為基線。

評估指標部分，連續型指標為均方誤差及皮爾森相關係數，均方誤差可評估實際值與預測值的誤差大小，誤差越小代表模型學習效果越佳，皮爾森相關係數可評估實際值與預測值之間的相關程度，避免演算法僅以特定值縮小誤差的未正確學習狀況。

我們接著將情感正負程度之連續值依據標記結果之平均值(5.3)離散化為偏向正面(5.3~9)及偏向負面(1~5.3)2 個類別，以離散型指標準確值(Accuracy)、精確值(Precision)、召回值(Recall)及 F 值進行評估，透過離散化指標我們可進一步觀察模型在不同類別的效能。準確值為全類別的效能評估，準確值的區間為 0 至 1，越高代表越準確；精確值、召回值及 F 值則為各類別的效能評估，區間亦為 0 至 1，精確值高代表預測類別與實際類別相符的比例高，召回值高則代表實際類別可被預測出來的比例高，而 F 值則是精確值與召回值的調和分數，F 值越高代表精確值與召回值的平均效能越高。

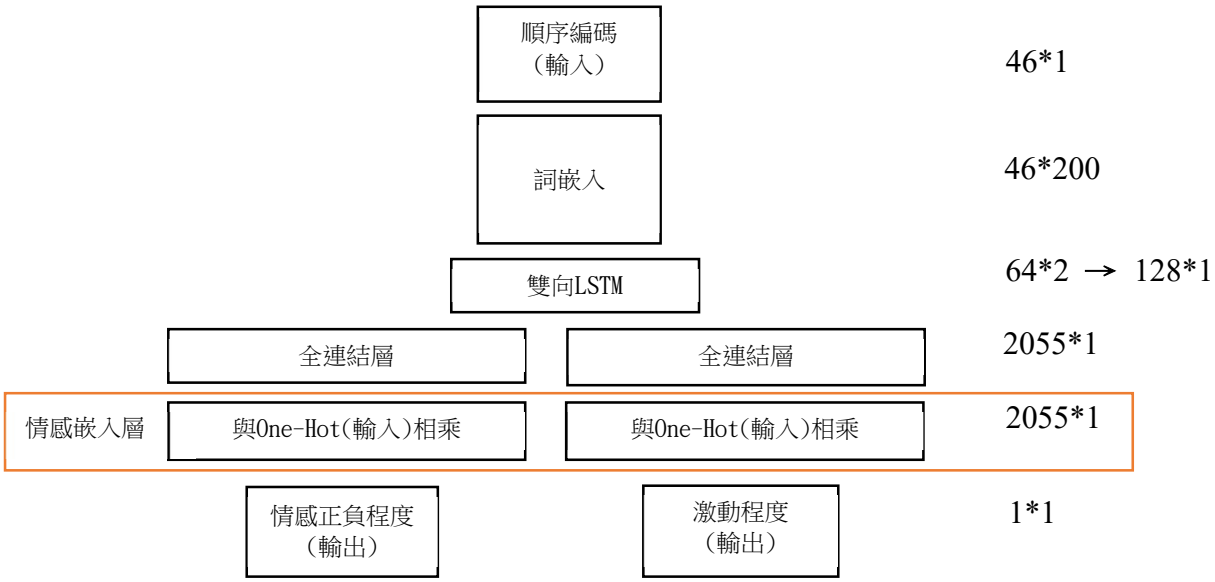
為了確認各模型是否能有效學習，我們第 1 組實驗先將全部的資料作為訓練集及測試集，各評估指標越高則代表學習效果越佳。而使用已知資料學習出來的模型預測已知資料，可能會有過度擬合的狀況發生，即模型過度合適特定資料導致無法用於其他資料，為了避免該狀況，第 2 組實驗進行 5 折交叉驗證，模擬測試集未被學習過的狀況，與第 1 組實驗相比，各模型效能通常會下降，若效能下降幅度過大即可能為過度擬合，導致模型應用在新資料的狀況較不理想。

在確定各模型的學習效能後，我們會希望用在新資料的預測上，但考量機器學習與深度學習皆為啟發式演算法，目的是針對因變數自動找出合適的自變數函數組合，因此處於一種難以被觀察的黑箱模式，不如字典法來的直觀及容易解釋。因此我們透過深度學習的嵌入技術，在學習過程中調整情感嵌入的隱藏層，定位每個詞對短句情感值的貢獻，再將隱藏層的權重取出後正規化(Normalization)，即為詞的維度型情感。透過前述嵌入技術取得新詞後，我們就可以評估新詞及自適應中文維度型情感詞典之效能，並與機器學習演算法效能進行比較，分析自適應方法是否有效。

本實驗當中所使用之深度學習架構如圖三所示，係將句子經過順序編碼後，使用預訓練的 200 維詞嵌入取得詞向量(46*200)，其中 46 代表句子的最大長度，經過長度為 64 的雙向長短期記憶(Long Short-Term Memory, LSTN)編碼後取得長度為 64*2 的向量，再將該向量連接成為 128*1 的向量作為情感程度及激動程度的共用編碼，解碼部分則使用情感正負程度及激動程度等 2 組全連結層，長度均為詞總數 2055*1，最後以長度為 1*1 的

全連結層作為輸出。另為定位詞的權重，取出新詞的模型會加上與句子輸入 One Hot 編碼相乘之隱藏層，長度同樣為 2055*1，使權重以詞存在與否進行調整。

前述深度學習模型的設計，在詞嵌入部分使用預訓練的詞嵌入，是因為本研究樣本數較少，在大量文本預訓練的詞嵌入可保留較好的詞位置資訊，而雙向長短期記憶則是透過由前往後及由後往前的編碼取得順序資訊，綜合兩者則為輸入的語意資訊編碼，此編碼器之設計在自然語言處理中已廣泛被使用。而解碼器部份我們假設輸入的情感正負程度及激動程度係每個詞貢獻而成的，因此使用全連結層乘上 One-Hot，定位每個詞的位置及貢獻程度，最後將貢獻程度正規化即為每個詞的維度型情感，此假設單純而易解釋。



圖三、自適應中文維度型情感詞典深度學習架構

四、實驗結果

在第 1 組實驗中，我們先以全部資料進行訓練及測試，確認資料是否可被有效學習及演算法是否正確學習，並以連續型指標及離散型指標評估學習結果。測試結果如表三所示，以極限梯度提升最佳，其次為支援向量機及深度學習，效能皆優於中文維度型情感詞典。

接著我們取出隱藏層，並將值正規化至原情感的尺度(Scale)，將其作為新詞的情感，透過字典法評估，新詞效能亦優於中文維度型情感詞典，因新詞係透過學習標記資料後所擷取出來的，而中文維度型情感詞典則是在通用領域取得，尚未適應至健保政策領域。

而考量新詞僅在本資料中訓練後擷取，為增加通用性，我們結合新詞與中文維度型情感

詞典為自適應情感詞典，僅加入尚未存在於中文維度型情感詞典的新詞，並去除單字詞，該詞典之評估結果顯示情感正負程度效能些許提升，說明中文維度型情感詞典在情感正負程度可能有補充效果。

而為了進一步觀察不同情感正負程度的效能，我們將值離散化至正面及負面，如表四所示，可觀察到新詞在負面效能高於正面，因此我們假設在本份資料負面較容易學習，並在後續進一步分析為何負面較容易學習。

在第 2 組實驗中，我們進行 5 折交叉驗證，確認演算法在學習時沒有過度擬合。結果如表五所示，以支援向量機最佳，極限梯度提升可能因過度擬合而使效能下降，而深度學習效能下降幅度最低，可能具有較佳的穩定性，3 種演算法雖效能下降但同樣優於中文維度型情感詞典。

表三、訓練測試連續型指標評估

演算法	均方誤差 (Valence)	相關係數 (Valence)	均方誤差 (Arousal)	相關係數 (Arousal)
極限梯度提升	0.0006	0.9998	0.0006	0.9996
支援向量機	0.3984	0.8953	0.2032	0.8513
深度學習	0.4650	0.8731	0.2518	0.8000
新詞	0.9593	0.7099	0.5180	0.4924
自適應中文維度型情感詞典	0.9503	0.7091	0.6209	0.3903
中文維度型情感詞典	1.4837	0.5628	1.4472	0.1454

表四、訓練測試離散型指標評估

演算法	準確值	情感正負	精確值	召回值	F值
極限梯度提升	0.9992	正面	1.0000	0.9984	0.9992
		負面	0.9983	1.0000	0.9991
支援向量機	0.9200	正面	0.9111	0.9322	0.9216
		負面	0.9294	0.9076	0.9184
深度學習	0.8817	正面	0.8546	0.9105	0.8817
		負面	0.9105	0.8546	0.8817
新詞	0.8400	正面	0.7900	0.8875	0.8359
		負面	0.8933	0.7997	0.8439
自適應中文維度型情感詞典	0.8083	正面	0.7351	0.8733	0.7982
		負面	0.8864	0.7585	0.8175
中文維度型情感詞典	0.7008	正面	0.6947	0.7167	0.7055
		負面	0.7074	0.6850	0.6960

表五、交叉驗證連續型指標評估

演算法	均方誤差 (Valence)	相關係數 (Valence)	均方誤差 (Arousal)	相關係數 (Arousal)
極限梯度提升	1.1807	0.6107	0.4990	0.4931
支援向量機	0.9656	0.6962	0.4546	0.5526
深度學習	0.9730	0.7017	0.5229	0.4987
新詞	1.4498	0.4825	0.6278	0.2429
自適應中文維度型情感詞典	0.9503	0.7103	0.6209	0.3895
中文維度型情感詞典	1.4837	0.5647	1.4472	0.1483

在離散化後，如表六所示，我們在本組實驗亦重複觀察到新詞負面效能高於正面效能，再次說明本實驗資料可能在負面較容易學習。另值得注意的是中文維度型情感詞典在正面的效能略優於負面，與演算法學習後的結果恰好相反，可能係因中文維度型情感詞典在健保政策領域中缺乏有效的負面詞。

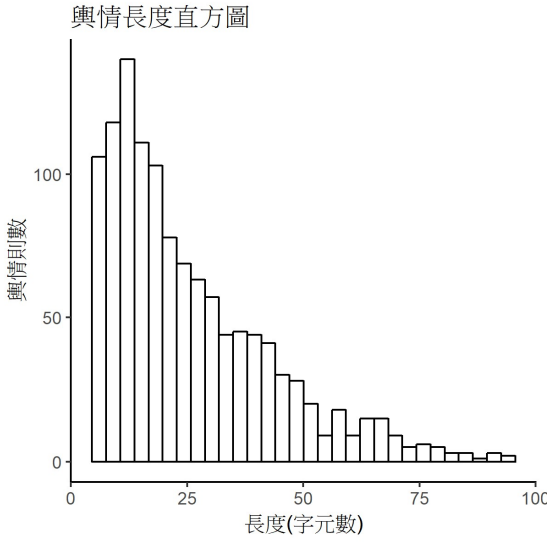
表六、交叉驗證離散型指標評估

演算法	準確值	情感正負	精確值	召回值	F值
極限梯度提升	0.7458	正面	0.7779	0.7422	0.7592
		負面	0.7119	0.7501	0.7300
支援向量機	0.7950	正面	0.7643	0.8238	0.7926
		負面	0.8263	0.7685	0.7961
深度學習	0.7992	正面	0.7701	0.8290	0.7966
		負面	0.8261	0.7755	0.7988
新詞	0.6875	正面	0.5613	0.7732	0.6444
		負面	0.8173	0.6399	0.7154
自適應中文維度型情感詞典	0.8083	正面	0.7353	0.8747	0.7979
		負面	0.8862	0.7579	0.8162
中文維度型情感詞典	0.7008	正面	0.6949	0.7170	0.7052
		負面	0.7082	0.6852	0.6959

字典法在效能一般難以超越機器學習及深度學習，因字典法每個詞只能對應一個值，即其表徵複雜程度較低，而機器學習及深度學習則是映射到高維空間的複雜函數，可處理的複雜程度並不相同。舉例來說，反諷詞在字典法仍為正面，即字典法無法分辨同一個詞在不同句子中的情況，但機器學習與深度學習則可透過句子中不同詞組成複雜的規則，區分詞在不同句子中的差異，即軟表徵(Soft Representation)可保留較多資訊量，使用詞嵌入取代詞袋模型即為其中一種應用。而經本實驗微調及適應後，自適應中文維度型情感詞典已貼近健保政策領域，在效能表現上略優於機器學習中效能最佳之支援向量機。

為了近一步釐清為何在負面學習效果較佳、正面學習效果較差，我們針對 1,200 則健保政策輿情進行統計及視覺化，分析資料的差異並檢視對學習結果的影響。首先是字元數的分布狀況，1,200 則健保政策輿情字元數之平均值為 25.73 個字元、標準差為 17.75 個字元，可從圖四中看出呈右偏態，資料集中在左側。

為了觀察字元少及字元多的差異，我們定義及篩選短輿情及長輿情，短輿情為平均值減標準差，即 7.98 字元以下，長輿情為平均值加標準差，即 43.48 字元以上，並觀察該 2 組輿情在正負面程度與激動程度的分布情況。從圖五中可得知短輿情在正面多於負面，而長輿情則相反，分析結果與常識相符：「滿意的人少回覆，不滿意的人多抱怨」。



圖四、健保政策輿情字元數分布

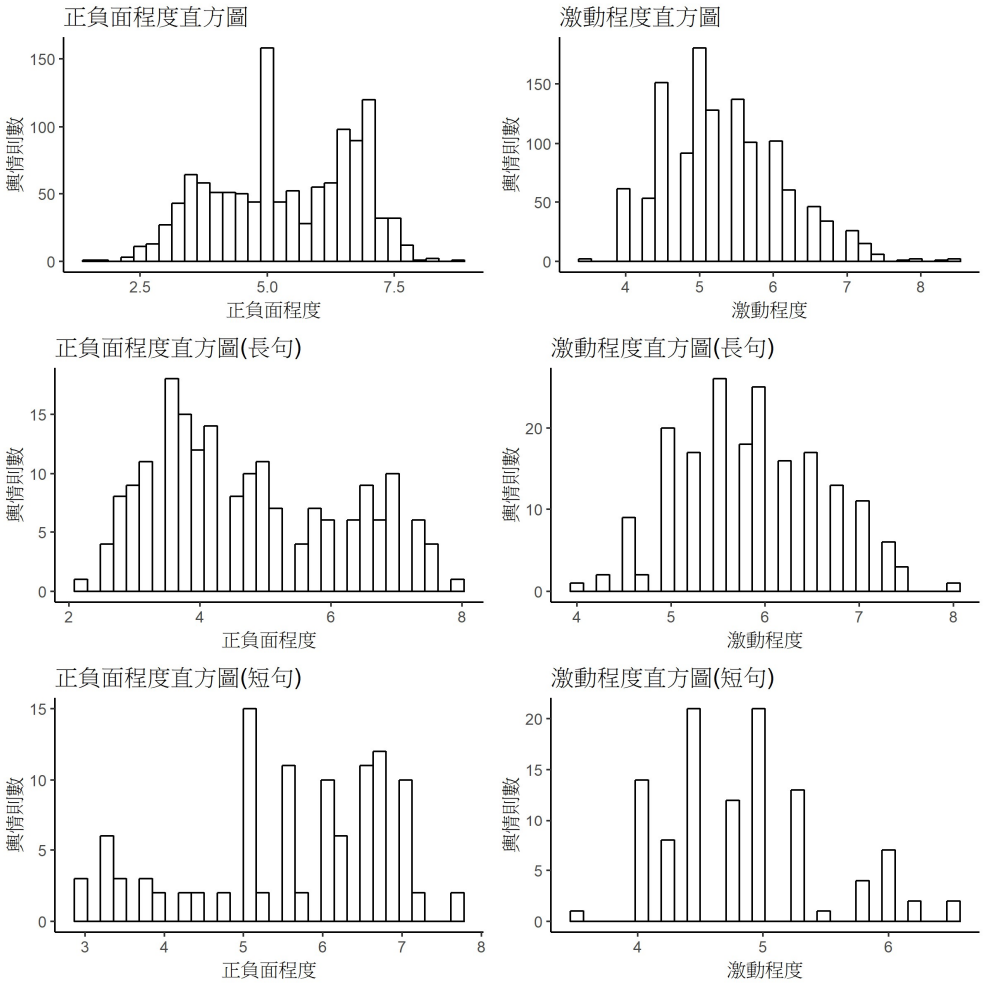
而短輿情偏向正面這樣的分布情形，即導致特徵不足，可能使學習的效果較差；反之長輿情多負向這樣的分布情形使特徵較多，則可能較適合學習。為解釋特徵數量的問題，我們取出短輿情最正面的 5 則及長輿情最負面的 5 則，可觀察出短輿情較易出現重複詞，長輿情之詞彙則較多樣。結合前述分析結果，當長輿情正面比例低但短輿情正面比例高的狀況同時發生時，可能是正面輿情不易被學習的原因。

表七、健保政策短輿情最正面 3 則

情感正負面程度	激動程度	字元數	輿情
7.75	6	5	很滿意健保
7.75	6.5	6	就很滿意啦
7.25	5.75	6	非常滿意健保

表八、健保政策長輿情最負面 3 則

情感正負面程度	激動程度	字元數	輿情內容
2.25	6.75	73	很怕健保會倒掉，有些人在A健保很過分，醫療系統不好，明明醫不好八九十歲中風還要氣切插管，浪費醫療資源，這樣很缺德，要規定八十歲以上不要有侵入性治療
2.5	6.75	72	健保藥物給付的條件越來越嚴苛，所以很多人會覺得一般診所的藥越來越差，寧願多花錢去醫院。鼓勵民眾先去診所治療，但又讓人感覺藥比較差，這樣怎麼會願意
2.5	6.75	60	醫療資源時常被濫用浪費，很多人明明不需要就診，還是跑來看醫生，還有很多中醫說要義診，叫老人家拿卡去刷，這都是浪費健保資源



圖五、健保政策長、短輿情情感正負面程度及激動程度分布

我們已知新詞在負面的學習效果優於正面，因此挑選 10 個較特別的負面詞如表九，可看出福利國、勞工、大陸人等特別的負面詞，在一般領域中則偏向中性，但在健保政策領域出現時可能較為負面，這些健保政策領域中特殊的負面詞可說明深度學習有學習到新詞的維度型情感，補充了中文維度型情感詞典在健保政策領域不足的部分。

相較傳統針對特殊詞個別標記之方法，從句子中自動提取新詞之方法已有效率的提升，惟本方法仍受資料數量及標記品質影響，當資料數量不足時或標記錯誤較多時，詞的學習效果就會不理想，因此所學習出之新詞並非完全正確，僅係根據基礎事實自動學習出來的結果，因此仍有正面之新詞夾雜負面詞或負面之新詞夾雜正面詞等錯誤結果。在自適應中文維度型情感詞典的方法下，前揭錯誤可透過增加資料數量或改善標記品質改善，若須確保新詞的正確性，則可透過人工篩選保留正確的新詞。

表九、經挑選後之負面新詞

新詞	情感正負面程度	激動程度
福利國	1.6754	6.9010
病房	1.6796	5.1799
倒掉	1.7034	5.0382
專利	1.7133	5.2003
費率	1.7259	5.0074
大陸人	1.7406	5.0936
勞工	1.7577	5.2355
掛號費	1.8044	3.1129
健保費	1.9870	5.0051
漲價	2.0198	4.9987

五、結論

本節實驗在正確性部分，我們蒐集了實際的健保輿情，並經過健保領域的研究人員進行情感標記後，比較中文維度型情感詞典與真實輿情標記結果，說明中文維度型情感詞典在健保政策領域效能較差的情形。

為了改進效能，我們透過機器學習及深度學習等演算法重新學習，並以連續型及離散型指標評估效能，在機器學習演算法中以支援向量機表現最佳，交叉驗證實驗之均方誤差為 0.97、皮爾森相關係數為 0.7，若不考慮黑箱問題，該模型可有效應用在情感預測。

而考量健保政策領域需要易理解的規則，因此我們提出透過深度學習隱藏層提取新詞的方法，即自適應中文維度型情感詞典，在經過效能評估後，驗證中文維度型情感詞典可透過增加新詞改善在健保政策領域的效能，其交叉驗證實驗之均方誤差為 0.95、皮爾森相關係數為 0.71，效能略優於支援向量機，且可直觀的觀察到每個詞的情感。

在資料分析的過程中，我們觀察到了正面輿情長度較短，而負面輿情長度較長的狀況，

符合「滿意的人少回覆，不滿意的人多抱怨」的常識，因此並非資料蒐集過度集中所導致。而正面輿情缺少特徵，可能造成學習效果略低於負面輿情，故在新詞提取中，負面詞有較佳的學習效果。

透過實驗我們初步驗證自適應方法，在正確性部分可以透過增加真實健保政策輿情資料量及標記人員數量等方式強化，在效能部分則可透過深度學習提取新詞進行自適應，而在解釋性部分則可透過人工篩選改進。綜上，本節研究之自適應中文維度型情感詞典，已可將中文維度型情感詞典適應至特定領域，強化情感分析的正確性，且效能在本實驗資料中優於機器學習或深度學習模型。

五、未來展望

本研究使用之深度學習模型仍有許多改進空間，例如：對於情感正負程度的貢獻未必每個詞都是同等重要，透過注意力機制可能會有更佳的效能；每個詞的情感嵌入到單一神經元僅取得詞的整體情感貢獻，但詞未必只有一種情感，在不同位置可能會代表不同情感，例如反諷時就常利用相同的詞表示負面情感，因此改為嵌入多個神經元或許可以取得詞在不同語境下的情感編碼，但更複雜的模型往往需要更大量且優質的標記資料。

因此考量標記任務為自適應中文維度型情感詞典重要的一環，當標記的數量或品質較差時，可能導致無法有效學習，因此為完善自適應中文維度型情感詞典的有效循環，應搭配有效的標記推薦方法。我們將在未來設計標記推薦系統，其透過隨機的刪減小樣本，監控深度學習模型的學習誤差，並結合刪減樣本數的懲罰係數設計目標函數，透過多次迭代使目標函數收斂，並將刪減的樣本即視為雜訊樣本，而雜訊樣本則假設可能為新詞數量較多或斷詞錯誤造成學習效果不佳，因此可透過修正結巴自定義詞典修正斷詞，再將雜訊樣本跟未被刪除的樣本進行詞彙差集，取得雜訊樣本中的差集詞彙，最後再透過搜尋引擎查詢差集詞彙取得較相似的新樣本，推薦領域專家進行標記，完善具方向性的標記及學習循環，降低標記成本，使自適應中文維度型情感詞典能更有效的持續發展。

六、致謝

本研究特別感謝陽明大學醫務管理所林寬佳教授、碩士生江婉琪及衛生福利部中央健康保險署相關人員，協助資料蒐集與標記並提供專業知識指導。本研究承蒙科技部 MOST 107 -2628-E-155-002-MY3 經費補助特此致謝。

參考文獻

[1] F. Z. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons," *Information Processing & Management*, vol. 56, no. 3, pp. 554-564, 2019.

[2] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 8, no. 4, 2018.

[3] J. Wang, L. C. Yu, K. R. Lai and X. Zhang, "Community-based Weighted Graph Model for Valence-Arousal Prediction of Affective Words," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 24, no. 11, pp. 1957-1968, 2016.

[4] L. C. Yu, J. Wang, K. R. Lai and X. Zhang, "Pipelined Neural Networks for Phrase-level Sentiment Intensity Prediction," *IEEE Transactions on Affective Computing*, 2018.

[5] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon based approach," *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, 2017, pp. 149-154.

[6] X. Fu, J. Yang, J. Li, M. Fang and H. Wang, "Lexicon-Enhanced LSTM With Attention for General Sentiment Analysis," *IEEE Access*, vol. 6, pp. 71884-71891, 2018.

[7] B. Shin, T. Lee, and J. D. Choi, "Lexicon Integrated CNN Models with Attention for Sentiment Analysis," *arXiv preprint arXiv:1610.06272*, 2016.

[8] Y. Zou, T. Gui, Q. Zhang and X. Huang, "A Lexicon-Based Supervised Attention Model for Neural Sentiment Analysis," *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 868-877, 2018.

[9] L. C. Yu, L. H. Lee, S. Hao, J. Wang, Y. He, J. Hu, K. R. Lai and X. Zhang, "Building Chinese affective resources in valence-arousal dimensions," *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 540-545, 2016.

[10] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," *IEEE Access*, vol. 7, pp. 43749-43762, 2019.

[11] P. T. Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 imagery," *Sensors,* vol. 18, no. 1, pp. 18, 2018.

[12] H. Dong, X. Hu, L. Wang and F. Pu, "Gaofen-3 PolSAR Image Classification via XGBoost and Polarimetric Spatial Information," *Sensors,* vol. 18, no. 2, pp. 611, 2018.

[13] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks,* vol. 61, pp. 85-117, 2015.

[14] M. Lin, Q. Chen and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[16] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167,* 2015.

[17] J. Wang, L. C. Yu, K. R. Lai and X. Zhang, "Tree-Structured Regional CNN-LSTM Model

for Dimensional Sentiment Analysis," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 28, no. 1, pp. 581-591, 2020.

[18] J. L. Wu, Y. He, L. C. Yu and K. R. Lai, "Identifying Emotion Labels from Psychiatric Social Texts Using a Bi-directional LSTM-CNN Model," *IEEE Access*, vol. 8, pp. 66638-66646, 2020.

[19] Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013.

[20] Y. Wang and S. Hu, "Exploiting high level feature for dynamic textures recognition," *Neurocomputing,* vol. 154, pp. 217-224, 2015.

[21] C. Yang, Z. Liu, D. Zhao, M. Sun and E. Chang, "Network representation learning with rich text information," *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[22] T. Mikolov , K. Chen, G. Corrado, J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:*1301.3781, 2013.

[23] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, 2013.

[24] L. C. Yu, J. Wang, K. R. Lai and X. Zhang, "Refining Word Embeddings Using Intensity Scores for Sentiment Analysis," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 26, no. 3, pp. 671-681, 2018.

[25] H. Keshavarz and M. S. Abadeh, "ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs," *Knowledge-Based Systems,* vol. 122, pp. 1-16, 2017.

[26] B. Shi, Z. Fu, L. Bing, W. Lam, "Learning domain-sensitive and sentiment-aware word embeddings," *arXiv preprint arXiv:1805.03801,* 2018.

[27] J. Barnes, R. Klinger and S. S. I. Walde, "Projecting embeddings for domain adaptation: Joint modeling of sentiment analysis in diverse domains," *arXiv preprint arXiv:1806.04381,* 2018.

[28] L. C. Yu, J. Wang, K. R. Lai and X. Zhang, "Refining Word Embeddings Using Intensity Scores for Sentiment Analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 671-681, 2018.

[29] X. Fu, Y. Wei, F. Xu, T. Wang, Y. Lu, J. Li and J. Z. Huang, "Semi-supervised aspect-level sentiment classification model based on variational autoencoder," *Knowledge-Based Systems,* vol. 171, pp. 81-92, 2019.

[30] S. Zhai and Z. Zhang. "Semisupervised autoencoder for sentiment analysis," *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[31] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, Y. Huang, "Semi-supervised dimensional sentiment analysis with variational autoencoder," *Knowledge-Based Systems,* vol. 165, pp. 30-39, 2019.