

國立台北大學資訊工程學系108學年度專題

應用自然語言處理與人工智慧技術 於常用刑事類別判決預測

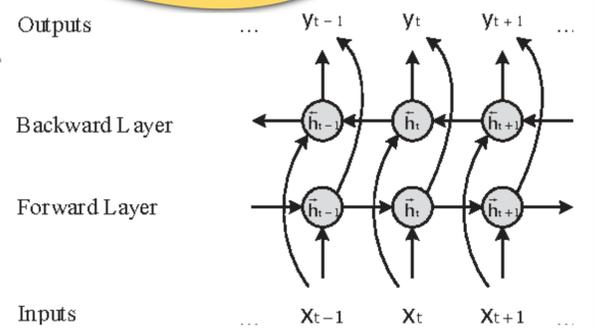
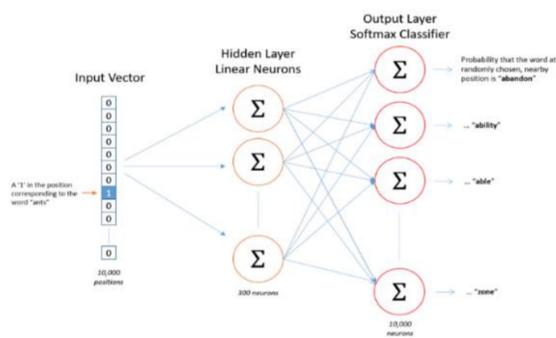


成員:葉宥麟 吳建宇 李侑程 黃品程

摘要

自然語言認知和理解是讓電腦把輸入的語言根據目的進行處理，本專題希望透過電腦系統的判斷，模擬法院刑事案例判決，給使用者參考，做為日後考量依據。

使用模型介紹



介面介紹



1. 打開網頁。
2. 按照右方案件範例的形式，在左方輸入所發生的事實。
3. 按下下方的提交並等待數秒。

Word2Vec模型

將每一個字詞轉換為以數值向量化形式的形式，將字詞向量化。一種方法為 Continuous Bag-of-Words，利用選定的詞周圍的詞來預測中間詞，一種方法為 Skip-Gram，會根據選定的中間詞來預測其周圍的詞。

BiLSTM模型

一般的LSTM模型只考慮到句子之中已出現的部分，但很多的時候句子，後半部的訊息也是重要的，因此BiLSTM會結合前向的LSTM和後向的LSTM，讓不管是句子中後半和前半的內容都考慮到。

結論與優點

1. 預先知道可能判決結果
2. 節省成本：時間、花費
3. 簡單化輸入：不需要專業法律用詞
4. 良好的準確率：

預測刑事種類

類別	傷害	竊盜	詐欺	毒品	總計
訓練數	29900	29900	29900	29900	119600
測試數	300	300	300	300	1200
正確數	297	391	297	297	1176
錯誤數	6	9	6	3	24
正確率	0.98	0.97	0.98	0.99	0.98

預測有罪無罪

類別	傷害	竊盜	詐欺	毒品	總計
訓練數	29900	29900	29900	29900	119600
測試數	300	300	300	300	1200
正確數	291	281	291	294	1158
錯誤數	9	18	9	6	42
正確率	0.97	0.94	0.97	0.98	0.965

預測刑期

類別	傷害(天)	竊盜(天)	詐欺(天)	毒品(年)
1	0~10 (54)	0~30 (57)	0~90 (18)	0~1 (66)
2	11~30 (60)	31~90 (66)	91~180 (30)	1~2 (42)
3	31~60 (33)	91~150 (42)	181~270 (36)	2~3 (27)
4	61~120 (21)	151~210 (27)	270~360 (69)	3~4 (33)
5	121~180 (42)	211~270 (33)	360~450 (57)	4~5 (27)
6	>181 (33)	>271 (15)	>451 (18)	>5 (48)
準確率	81%	82%	76%	81%

