10994015 資管碩二 李承諺

心得:

本次我採用子霆的以深度學習技術進行消息驗證-以MSCI台灣指數為例，

深度學習一直都是有趣的領域而此領域又可以應用在許多不同的地方像是股市分析，而子霆也剛好在做股市的消息分析，他的題目很明顯就是要透過這個分析來賺錢，以MSCI的資料來去分析股市接著筆其他人先行布局，但這點就是透過人家的分析產生的資料，若用人加以分析過的資料來去分析，是不是會跟原本的分析的依樣而變得沒有意義，或是子霆能發現什麼其他人未發現的資訊來分析股市，趁機賺一波?我也很期待她之後的論文研究結果。

相關論文:

**使用深度學習進行基於影片的台灣手語辨識**

(1)基本資料:

黃明翰（2021）。使用深度學習進行基於影片的台灣手語辨識。國立陽明交通大學數據科學與工程研究所

(2)研究問題:

因應多元化社會，現今手語已成為了一種語言，而手語在地域上也會有不同的手語意義，而目前台灣手語翻系統較為少見，因此此研究針對台灣手語來進行手語辨識，而手語翻譯的工作也成為了標準配置，但現在的手語翻譯通常需要一個人在旁翻譯，或是使用穿戴式設備或是深度攝影機來進行手語辨識，但此方法實在耗費人力資源以及不具備方便性與普遍性，也需要耗費龐大的金錢來購買設備，非常不適合日常使用，因此此研究要使用最新的電腦視覺與深度學習技術來進行手語辨識，使用權人體姿態估計的Darkpose抽取人體的骨架資訊，再結合神經網路模型將骨架資料丟進去進行手語辨識，最後將此模型應用到便攜式的輕量級設備如手機webcam等，解決手語辨識需要人力資源以及耗費設備的問題。

(3)研究方法/提出之方法論

此研究為實驗法

此研究提出一個台灣手語辨識系統來對台灣的手語進行辨識，因為此研究主要解決能及時幫助聾啞人士進行無障礙溝通及聾啞人士遇到緊急情況需要及時求助或是幫忙的狀況，就基於以上兩點，此研究歸類出了四種手語分別為感受類、求救類、溝通類、及日常類以符合此任務的需求，而此模型輸入了台灣手語影片，並輸出40種結果，拍的每部影片皆為1920\*1080畫素，幀率為30幀，拍攝角度皆為手語者正面，即使稍有偏斜也與手語的規定範圍相符。總計有4男2女比出手與，年齡為19~45歲，影片包含7種不同背景，整理後影片為746部，皆著再將影片用3D-CNN轉為RGB圖片，每張照片將畫素縮小至256\*256以方便訓練，經過轉換後得到59610張原始影片做為RGB資料集。

接著再將資料集4:1切分為訓練集以及驗證集，最後訓練集影片有619筆，驗證集則有127筆。

此模型會分別計算Top-1、Top-3、Top-5的機率，首先模型先用市面上已成熟的全人體估計模型DarkPose來抓取骨架資料，分別為抓取全關鍵點以及特徵關鍵點的骨架資訊，接著再用Yan et al.提出的GCN模型作為手語的辨識，透過分析骨架的姿態來辨識手語意義，最後進行模型集成，最終會獲得人體姿態模型以及RGB影像輸出，兩模型皆輸出長度40的向量，分別代表40種手語的機率，最後將兩種預測結果不同德銓重加總得到更精準的預測結果。

(4)研究結果:

第一階段實驗全關鍵點(133個特徵)去除掉下半身後的121個關鍵點進行訓練集預測，實驗結果Top-1需要60個epoch他的loss才趨於穩定，準確率為94.90，Top-3、Top-5則在20個epoch就趨於穩定，準確率皆為99.32%。

根據混淆矩陣顯示，其中立刻、不是、謝謝為一組表現較差的分類，推測是因為太多雜訊分散了手勢部分準確率，另外不知道、厭惡也表現較差，這兩種在表情上非常相似，但手勢上卻明顯不同，從這兩種表現較差的大類可看出有牽涉到表情變化的手語影響較大，因此在模型的表現不佳。

第二階段為特徵關鍵點，抓取出前面121個特徵中的39個足以代表人體關鍵部位的特徵關鍵點來做訓練和測試，觀察去除多餘資訊時是否會表現較佳，此實驗進行100個epoch結果為:

Top-3、Top-5的在20個epoch穩定下來準確率為95%，而Top-1在40個epoch就穩定下來比全關鍵點還快，此實驗Top-1準確率為97.96%，Top-3、Top-5準確率為100%，並觀察混淆矩陣可看出在第一階段較差的不知道和厭惡在此階段表現較好，而立刻、不是、謝謝也在此階段能順利的區分出來，因此此模型更具焦在手勢上的變化。

第三階段使用RGB圖形直接辨識，將59610筆影像資訊訓練100epoch，實驗結果為在70個epoch趨於穩定準確率達到95%，Top-1準確率為97.62%，Top-3、Top-5準確率為99.32%。由混淆矩陣看出謝謝、不知道是較差的分類，可看出在有表情變化上的手語表現很不穩定，但在手部變化較小的手語辨識上表現得比前兩階段還要更好，例如立刻、不是、謝謝等分類。

第四階段為模型集成，將訓練完的姿態模型及RGB模型集成看看是否能得到更好的結果，而兩模型權重為45時效能最佳，而集成後Top-1準確率為98.64%，Top-3、Top-5皆為100%。根據混淆矩陣看出，預測錯誤的剩下不知道、謝謝、不是三個分類，但從前面階段可看出遇到動作類似但表情不同的手語都被改善了。