**自然語言處理導論期末报告**

**Deep Learning for Event-Driven Stock Prediction**

**第五组 组员：A10515001 李大祥 A10515002 張天佑 A10515003 鄧鵬宇**

现在使用计算机技术去预测气象灾害以及用户的购买意向已经非常常见了，也有很多成熟的做法，我们这次实作的题目使用深度学习技术预测股票涨跌还算是一个新鲜的课题。在金融领域，很多年来都是由人类去主导的，虽然借助各种数学模型以及演算法，但是依然无法脱离人的选择，这也是为什么华尔街的分析师为何薪水那么高的主因。我们参考的论文也是从此下手去证明使用深度学习可以在某种程度上预测股票的涨跌。

影响股票的涨跌有着多方面因素，比如公司情况，政府决策，历史股价，法律变动，亦或是市场的情绪。基于这些，论文中主要是采取从新闻中结构化的事件来表示。我们在讨论实作上也提出了基于新闻的语义情感分析去判断，像是积极的新闻比如某公司财报利好利润上升等，亦或是赤字之类的负面情报。最后我们决定采用和论文类似的方式，从新闻标题上着手，使用OpenIE去抽取我们需要的资讯，学习出一个通用的关系抽取模型。这里使用的是ReVerb的工具帮我们把新闻标题分割成O1，P，O2的格式，举例来说“44 African countries agree free trade agreement, Nigeria yet to sign.”这个新闻的标题会被分成O1: african countries，O2: free trade agreement，P: agree的形式，类似从语法角度找出主谓宾的结构。

網絡模型和baseline的相同點就是都是基於金融新聞，訓練資料和測試資料也都相同，關於訓練集和測試集也找到一些現有的，不過我們想的是用一些新聞的API自己去抓取。不同點主要是采用event embedding和deep learning CNN技術。實驗方式是如果模型指出第二天個股價格會上漲，那麼虛構交易者將以開盤價格買入10000美元。購買後，交易員將持有該股票一天。在持有時間內，如果股票可以獲利2％或更多，交易者立即賣出。否則以收盤價賣出股票。最後實驗結果是，event embedding對於股市預測是有效果的，能夠更好的預測特徵，結果也是比較好的。

我們自己的實作采用的方式是LSTM，考慮了時間信息的RNN技術，最後我們實驗結果也逼近了論文中CNN的方法，結果也能夠證明我們的方法也是有效果的。