國立高雄科技大學
企業管理系碩士班
碩士論文

應用語言模型於線上招募詐欺偵測

An Application of Language Model on Online Recruitment Fraud Detection

研 究 生：洪英皓

指導教授：余銘忠　博士

中 華 民 國 110 年 6 月

應用語言模型於線上招募詐欺偵測

An Application of Language Model on Online Recruitment Fraud Detection

研 究 生：洪英皓

指導教授：余銘忠　博士

國立高雄科技大學

企業管理系碩士班

碩士論文

A Thesis
Submitted to
Department of Business Administration
National Kaohsiung University of Science and Technology
In Partial Fulfillment of Requirements
For the Degree of Master of Business Administration

June 2021
Kaohsiung, Taiwan, Republic of China

中華民國 110 年 6 月

應用語言模型於線上招募詐欺偵測

學生：洪英皓　　指導教授：余銘忠　博士

國立高雄科技大學　　國立高雄科技大學

企業管理系碩士班　　企業管理系　教授

國立高雄科技大學企業管理系碩士班

摘要

隨著網路科技的快速發展與進步，眾多生活型態也隨之改變。在人才招募方面，也逐漸從傳統的平面媒體轉成網路媒體來進行招募。雖然改善過去在招募人力時的成本，卻導致衍生許多網路求職陷阱，造成不少民眾受害。儘管過去政府與人力招募平台大幅的宣導詐騙防範理念，但詐騙手法日新月異的變化，改善的幅度非常有限。

近年來人工智慧的技術快速發展，在不少的領域中皆有許多的突破。而本研究發現過去針對於網路招募詐欺偵測議題，較少研究提出深度學習的方法來應對。因此本研究基於語言模型(BERT, Word Embedding, Topic Model)來提出兩種模型架構(BERT-DNN, GRU-DNN)，藉此來探索適合應用在網路招募詐欺偵測的語言模型為何。

本研究以 EMSCAD 作為模型訓練與測試資料集，並使用維基百科的語料庫作為預訓練詞向量模型之語料庫，並交互實驗以及比較不同的詞向量模型(fastText, word2vec)對所提出的模型架構之影響力。在研究結果中，發現以 RNN 為基礎的模型總體表現皆優於以 BERT 為基礎的模型，兩者在平衡後的準確率方面分別取得 97.33%與 94.33%，在 f1_score 則是 77.13%與 59.76%。此結果也優於過去研究採用的機器學習模型表現，因此未來若使用本研究所提出的模型架構進行偵測，能有效的降低民眾受害之機率。

關鍵詞：詐欺偵測；主題模型；自然語言處理；深度學習；BERT

An Application of Language Model on Online Recruitment Fraud Detection

Student：Ying-Hao Hung　　Advisors：Dr. Min-Chun Yu

Department of Business Administration
National Kaohsiung University of Science and Technology

ABSTRACT

As the information technology rapidly developing, many people's lifestyles have also changed. As for recruitment, it has shifted from traditional offline media to online media. Even though online recruitment can decrease labor cost, it has brought on online job scam that has many applicants suffered. The government and job banks have disseminated knowledge about anti-fraud to decrease the number of victims in the past. However, scammers are getting more sophisticated and coming up with new ways to scam job seeker, hence many people still suffer.

The technology of artificial intelligence has been developed rapidly in recent years, so there have been many breakthroughs in many fields. Nevertheless, not many researches are using deep learning method to deal with the issue of online recruitment fraud detection. Therefore, this study proposed two model architectures (BERT-DNN, GRU-DNN) based on language model (BERT, Word Embedding, Topic Model) to develop a language model that is suitable for online recruitment fraud detection.

In this study, we used EMSCAD dataset to train and test our model. In addition, Wikipedia corpus is utilized to pre-train word embedding models (fastText, word2vec) to conduct a series of experiments to find out the best word embed model for our proposed model. The results show that the performance of GRU-DNN is better than BERT-DNN in each experiment. The balance accuracy for GRU-DNN and BERT-DNN is 97.33% and 94.33%; f1_score is 77.13% and 59.76% respectively. Therefore, the proposed model performs better than the conventional machine learning model. Hence, it is suggested that the proposed model can effectively reduce the probability of job applicants being scammed if implemented for online recruitment detection.

Key Words：Word Embedding, BERT, Fraud detection, Topic Model, Natural Language Processing