

摘要

近年來因少子化衝擊，各大學接受申請入學考試的名額也隨之減少，各大學之入學審查單位傾向以書審來評估申請入學的學生，學校提供一個尺規供審查委員替審查資料評分，為了減輕審查委員的負擔，需要對眾多格式繁雜的書審資料去評分，以及不同審查委員各自的主觀想法導致評分結果不公平，因此，本研究提出自動評估學生入學資料之審查方法，將審查資料配對尺規量表並得到分數，本方法的主要貢獻可以減輕人工審查的繁瑣程序，並且提供一個較客觀的結果，消除審查委員的個人主觀意見的影響，來達到有效率與公平性的評估方法。本研究主要方法步驟如下：(1) 替尺規制定知識圖 (2) 提取審查資料特徵 (3) 配對尺規與審查資料 (4) 提供結果給審查委員參考，本研究提出的方法可以用來自動依據尺規評估審查資料，並得到相對應的分數。

關鍵字： LaTeX、中文、論文、模板



Abstract

Abstract

Keywords: LaTeX, CJK, Thesis, Template



Contents


	Page
摘要	1
Abstract	2
Contents	3
Chapter 1 諸論	1
Chapter 2 相關研究	4
2.1 人才招聘分析 (Recruitment Analysis)	4
2.2 文本挖掘 (Text Mining)	5
Chapter 3 背景	7
3.1 CKIP 中文斷詞	7
Chapter 4 問題描述	8
Chapter 5 結論	9
References	10



Chapter 1 諸論

在升學階段，許多大學開放學測個人申請、甄試入學申請等等的申請入學方式，在推甄名額越來越多，考試名額越來越少的情況下，學校越來越傾向以書面審查當作初試的審查標準。然而，現在的選才，除了課業表現外，社團活動、學習特質、動機與能力更是重要，學校希望能看到學生的更多面向，而不單單只是成績上的數字，審查資料時，更需要看見學生自發性學習的精神、學習動機和人格特質，書面審查資料屬於半結構化資料，包含了結構化資料的部分如：姓名、生日、住址等等，以及非結構化資料如：自傳、就讀動機等等，由於每個學生的書面資料格式不同，審查委員難以有效地找到資料中重要的訊息。


審查委員依據尺規評估學生的審查資料是否適合本科系，審查資料包含了學生的基本資料、就讀高中、成績、技能、專長、興趣、比賽經驗、幹部經驗、社團活動、自傳、讀書計畫、就讀動機等資料，學校提供一個尺規，其中包含了學生的一般學習表現、資工相關學習表現、多元表現、自我學習能力，每一項有15級的評分，其中一般學習表現根據學生修習數學、物理與英文課程的表現來評分，表現分別分為傑出、優秀、中上表現、中等表現及表現不佳。多元表現根據參加程式相關的競賽或課程，參與國際性、地區性或是校內性的數學、物理或是資訊科技等比賽，以及相關程式碼作品的表現及社團幹部經驗去評分。自我學習能力則是依照讀書計畫的縝密度以及就讀動機去評估學生。審查委員會根據這個尺規對書面資料打分數，最後以這個標準去評估學生是否可以錄取，然而在審查學生



的書面資料過程中，可能會面臨：(1) 審查委員分配到一部份的書面資料，由於各個審查委員的主觀意識不同，導致評分標準有落差，對審查有失公允，且對學生也不公平。(2) 審查資料眾多，加上資料格式繁雜，審查委員無法仔細審查資料。(3) 尺規的定義模糊，造成審查委員難以判斷(4) 人工審查費時費力為了減輕審查委員的負擔，在眾多格式繁雜的書審資料中找到學生的特質，所以本研究提出自動評估學生審查資料的方法，希望能透過深度學習的方法讓機器能夠自動配對審查資料中的文字到尺規，並且得出分數，因此，我們提出了自動評估學生審查資料的方法，能夠幫助審查委員在審查書面資料上的效率與公平性，從大篇幅的書面資料中，提取書面資料中有用的資訊如：學生的競賽表現、學業成績、幹部經驗與就讀動機等等，來評估這個學生是不是適合某個科系，如何利用科技有效地從審查資料中找到合適的學生，將是我們面臨到的課題。

隨著人工智慧的普及，在人才招聘分析 (Recruitment Analysis) 中，許多文獻利用人工智慧的技術來做履歷匹配、提取的應用，他們著重分析面試者的技能是否符合公司所需的技能、條件，近年來，隨著求職網站不斷蓬勃發展，在成千上萬的工作及履歷中，對於公司如何在這麼多的履歷中找到適合自己的員工，以及求職者如何在這麼多的工作中找到適合自己的，因此有許多人開始研究個人-工作配對問題 (person-job fit)，在相關研究中對於文本的表示，沒有考慮到不同用詞但代表相同意義的用語，如具有軟體開發經驗跟寫過 android app，然而大部分的研究只針對簡歷中的一部分資料去配對如：工作經驗，無考慮到整份履歷中的其他資料，如：教育程度、技能及自傳等等的配對。大部分的研究都是簡歷與工作需求這樣短文本—短文本的配對，缺少長文本—短文本的配對研究。

在我們的研究中審查資料屬於半結構化的長文本資料，表格包含畢業學校、專長、技能、幹部經驗等等的資料，以及自傳、讀書計畫、申請動機等長文本資料，將審查資料轉換成好的表示將是一大挑戰，尤其長文本在轉換特徵表示很難



捕捉到語義所以導致配對效果不好，我們提出的自動評估學生審查資料的方法，能夠進行長文本—短文本的配對，配對審查資料跟尺規。尺規是一個短文本，尺規是一個表格，而這個尺規是很廣泛且模糊的標準，所以我們要先對這個尺規建立一套方法，像是國際性比賽、有程式設計經驗且表現傑出，何謂國際性比賽、程式設計經驗，我們要如何去配對這些詞語，以便機器了解這個尺規並且能夠根據尺規去評分，所以我們運用建立知識圖的方法，替尺規去擴增豐富度。然而語言的表達形式很多，尺規中的要求實際上審查資料也滿足這個要求，但是表達方式或用詞用語不一樣，這時候透過知識圖得到表示會包含相關資訊，不但有利於配對，在消弭不同用語用詞也有幫助。


本研究設計了評估學生審查資料方法，利用配對審查資料與尺規去量化審查資料的分數，首先，對尺規資料去建立知識圖，將審查資料利用結合知識圖的文件編碼器 (Document Encoder) 進行編碼得到知識表示 (Knowledge Representation)，對於尺規中有程度上差別的評分如：傑出、優良、普通等等，制定一個方法給予審查資料尺規程度嵌入 (Rule-based representation)，接著，拼接文本表示及尺規程度嵌入得到增強表示 (Enhanced Representation)，將增強表示送到配對模型 (Match Model) 並與尺規配對得到最終的審查資料分數，透過機器自動對審查資料評分，能夠提供審查委員一個較客觀的參考。



Chapter 2 相關研究

2.1 人才招募分析 (Recruitment Analysis)


Chou, Yi-Chi, and Han-Yen Yu. [3] 提取履歷中的特徵，使用歸一化公式計算提取特徵的分數，這些分數分別對五個領域的工作中計算技能、經歷以及特質的分數，還有 DISC 的分析，並且使用 TF-IDF 的公式計算簡歷與職缺內容的相關度來推薦相關職缺給求職者。Maheshwary, Saket, and Hemant Misra. [6] 在孿生網路 (Siamese Network) 用使用卷積神經網路，工作要求及履歷為輸入，計算輸入之間的相似度，但他們忽略了句子間的語義資訊。Qin, Chuan, *et al.* [7] 提出一個基於循環神經網路 (Recurrent neural network, RNN) 的方法為 Basic Person-Job Fit Neural Network (BPJFNN)，用兩個雙向長短期記憶 (Bi-directional Long Short-Term Memory, BiLSTM) 得到工作經驗及工作要求的語義表示，接著將語義表示輸入到一層的類神經網路中，得到配對分數，有鑑於在履歷或工作要求中，不同的詞語在不同的位置有不同的重要性，Qin, Chuan, *et al.* [7] 在 BPJFNN 上加入了注意力機制 (Attention)，根據不同詞、不同句子之間的重要性得到工作要求及履歷工作經驗的表示，以及一條工作需求與各條工作經驗的相關度表示。接著用這些表示來預測配對分數。Zhu, Chen, *et al.* [8] 提出 Person-Job Fit Neural Network (PJFNN) 能將求職者過往的工作經驗來配對職位需求，工作需求及履歷中的工作經驗利用卷積神經網路 (Convolutional Neural Networks, CNN) 得到投影到同



一個空間的表示，接著計算在這個空間中的工作需求表示與工作經驗表示之間的距離，這個距離就是他們的相似度，來當作他們是否配對的依據，他們只使用履歷中一部分的資料作為預測，對於。由於求職是一個雙向的過程，雙方的意願是很重要的，所以 Le, Ran, *et al.* [4] 提出 Interpretable Person-Job Fitting(IPJF) 將雇主與求職者的意願加入模型中，首先預測招募者對求職者的意願以及求職者對此工作的意願，利用預測雙邊意願的過程中產生的隱藏特徵來預測配對的機率。Bian, Shuqing, *et al.* [2] 應用遷移式學習（Transformer Learning）中領域自適應（Domain Adaptation）的方法在工作配對的問題，用結構對應學習演算法（Structural Correspondence Learning，SCL）得到遷移過後的工作要求及履歷表示，在計算配對表示（match representation）的過程也是可以遷移的，最後經過多層感知器（Multilayer Perceptron，MLP）來預測最終的配對結果，此方法可以解決樣本不充分的工作領域的配對問題。Bian, Shuqing, *et al.* [1] 除了單純對文本做配對之外，還做了一個基於關係的配對模組（Relation-based Matching Component），某個工作要求與其他工作要求相似程度很高時，那麼已經與其他工作要求配對到的履歷應該也會與這個工作要求相似，在訓練配對時將純文本配對模型加上關係的配對模組進行預測工作與履歷的配對程度。

2.2 文本挖掘（Text Mining）

傳統的方法將文本表示為詞袋（Bag-of-words），統計每個詞出現的次數。詞頻逆文檔頻率（Term frequency-inverse document frequency，TF-IDF）是為了解決詞袋無法分辨常用詞以及不同詞語對文本的重要性問題，TF-IDF 可以過濾掉常見的及無關緊要的詞語，賦予關鍵字比較高的權重，詞袋跟 TF-IDF 都忽略詞語的順序。詞嵌入（Word Embedding）將詞對應到向量中的維度，將句子中的字轉成向量表示，並且考慮了句子中詞的順序，當字詞過多向量會變得很龐大。



Word2Vec 考慮到上下文，語意相似的詞有較近的距離，但是只看周圍幾個詞，詞向量資訊量不足，Doc2Vec 考慮詞序後算出代表一語句段落的向量。以上提到的都是靜態詞向量，無法解決一字多義的問題。ELMO (Embeddings from Language Models) 可以解決同義詞的問題，每個詞向量是雙向語言模型不同層的資訊，能夠捕捉詞義與上下文的資訊，但是這兩個方向的模型其實是分開訓練的，只是在最後做了個簡單相加，導致在單個方向看不到另一個方向的詞，有時候句子中的字同時依賴左右兩個方向的某些詞。基於變換器的雙向編碼器表示技術 (Bidirectional Encoder Representations from Transformers, BERT) 使用雙向變壓器 (Transformer)，使用遮罩的預測方式可以理解雙向上下文的能力，而非單個方向。

Liu, Bang, *et al.* [5] 提出長文本的配對方法，將文本轉換成圖 Concept Interaction Graph (CIG)，使用關鍵字提取演算法提取關鍵字，每個關鍵字為一個節點—概念 (Concept)，每個句子會附加到相關的概念中，先做局部的配對即對概念中的句子學習每個節點的配對向量 (matching vector)，然後使用卷積神經網路提取配對向量的特徵，最後使用多層感知器 (Multilayer Perceptron, MLP) 在此特徵預測上預測兩個長文本的是否相似。



Chapter 3 背景

3.1 CKIP 中文斷詞

對於繁體中文很好的效果，也較貼近台灣人的用詞用語，具有指代解、專有名詞辨識等功能，使用 CKIP 幫助我們在審查資料上進行斷詞。



Chapter 4 問題描述

我們定義一個備審為 D ，尺規為 R ，其中備審由 i 個句子組成，其中每一個句子表示成 d_i ，備審可以表示成 $D = \{d_1, d_2, \dots, d_i\}$ ，尺規由 j 條規則組成，其中每一條規則為 r_j ，尺規可以表示成 $R = \{r_1, r_2, \dots, r_j\}$ ，審查資料經過結合知識圖得到的表示法 (Knowledge Representation) 為 KR ，根據尺規得到的審查資料尺規程度嵌入 (Rule-based level representation) 為 RL ，拼接 KR 與 RL 得到增強表示 (Enhanced Representation) 為 ER 。最後經過配對模型得到配對分數 S 。



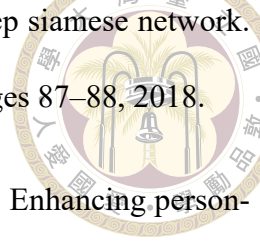
Chapter 5 結論

我們的研究可以自動評估審查資料，將審查資料配對到尺規中，並得到分數，先前的研究少有長文本到短文本的配對，但是在我們的研究中可以有很好的配對效果，透過建立知識圖來取得包含更豐富的表示，讓配對效果更好。自動評估審查資料可以提供審查委員一個參考，減少人工審查費時費力的工作，以及提供較客觀的參考，消弭審查委員間不同主觀意見的影響，而且也有良好的配對結果，並且得到分數。



References

- [1] S. Bian, X. Chen, W. X. Zhao, K. Zhou, Y. Hou, Y. Song, T. Zhang, and J.-R. Wen. Learning to match jobs with resumes from sparse interaction data using multi-view co-teaching network. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 65–74, 2020.
- [2] S. Bian, W. X. Zhao, Y. Song, T. Zhang, and J.-R. Wen. Domain adaptation for person-job fit with transferable deep global match network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4812–4822, 2019.
- [3] Y.-C. Chou and H.-Y. Yu. Based on the application of ai technology in resume analysis and job recommendation. In *2020 IEEE International Conference on Computational Electromagnetics (ICCEM)*, pages 291–296. IEEE, 2020.
- [4] R. Le, W. Hu, Y. Song, T. Zhang, D. Zhao, and R. Yan. Towards effective and interpretable person-job fitting. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1883–1892, 2019.
- [5] B. Liu, D. Niu, H. Wei, J. Lin, Y. He, K. Lai, and Y. Xu. Matching article pairs with graphical decomposition and convolutions. *arXiv preprint arXiv:1802.07459*, 2018.

- 
- [6] S. Maheshwary and H. Misra. Matching resumes to jobs via deep siamese network. In *Companion Proceedings of the The Web Conference 2018*, pages 87–88, 2018.
- [7] C. Qin, H. Zhu, T. Xu, C. Zhu, L. Jiang, E. Chen, and H. Xiong. Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pages 25–34, 2018.
- [8] C. Zhu, H. Zhu, H. Xiong, C. Ma, F. Xie, P. Ding, and P. Li. Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Transactions on Management Information Systems (TMIS)*, 9(3):1–17, 2018.