



台灣人工智慧學校

新聞推薦系統

News recommendation systems

學員：蔡堯力 (Eric Tsai)

Executive Summary:

本文結合機器學習和神經網路的演算法，透過平台上使用者的歷史瀏覽紀錄和新聞內容推薦使用者感興趣的新聞。使用的演算法包括 k-means、KNN 和 Bert 模型(2018)，分析過程也有應用自然語言處理(NLP)相關的技術，如 average word vectors、Doc2Vec、bag of words、TF-IDF 和 ELMo 等將文章向量化的方法。

新聞的主要目的之一就是傳遞資訊，讓社會大眾能快速了解自己生活圈以外的事，而推薦系統就是希望在資訊繁雜且選擇多樣的情況，能準確的提供使用者感興趣的選擇標的。因此本文將推薦系統應用在財經新聞平台(MoneyDJ 理財網)上。我們主要將推薦方法分為熱門度推薦、相似使用者推薦和個人化推薦。

- 熱門度推薦：透過新聞標題與內文預測新聞會不會變熱門，來優先曝光最新且模型判定會變熱門的新聞。
- 相似使用者推薦：透過過去使用者觀看新聞的產業標籤統計資料進行分群，並在不同群中找出點擊紀錄相近的使用者，找到目標使用者沒有看過，但相似使用者有看過的新聞進行推薦。
- 個人化推薦：透過 NLP 的技術將使用者過去看過的新聞轉成向量，並當作使用者的新聞偏好，最後透過向量的相似性，找出與使用者偏好相近的新聞。

本文目的在於透過圖示的方式解釋如何應用人工智慧的技術，了解使用者在平台上的偏好和特性，以達到自動化精準推薦新聞的目的，下一頁展示了新聞推薦系統的推薦流程和當中應用的技術。

建構推薦系統的資料集合包含使用者行為資料(User_Actions)、新聞資料集(News_info)和登入環境(ENV)，其相關變數列於下表當中。

Dataset

Datasets		
	Variables	Records
ENV	user_id (7,493,723)、browser (65)、OS (15)	11,159,911
News_Info (1999 ~ 2018)	news_id、date、news_content、news_title、news_industry (895)	48,503
User_Actions	user_id (7,419,810)、news_id (605,144)、click_time	13,170,863

新聞推薦系統

新聞點擊資訊

user_id, news_id, click_time
date, news_content, news_title, news_industry



對使用者做
產業貼標

k-means 使用者
分群

在第 N 群使用者中，
用 KNN 找出點擊紀錄
最相似的 K 個使用者

群組使用者推薦

News 1
News 2
News 3
⋮

文章向量化

計算使用者新聞
偏好向量

用 cosine similarity 計
算新聞與使用者偏好的
相似性

個人化新聞推薦

News 4
News 5
News 6
⋮

熱門推薦新聞

News 7
News 8
News 9
⋮

Word Embedding

1. average word vectors
2. Doc2Vec
3. bag of words
4. TF-IDF
5. ELMo

分不同產業，再根據點擊
數量定義產業內熱門新聞

BERT 文本分類模型