



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。使用者受利於推薦系統的原因有很多，主要可分為熱門度推薦、相似性推薦和主動性推薦。

- 熱門度推薦：透過熱門度推薦可以讓使用者更快速的了解其他人在關注什麼，以獲得一般性的資訊和認知。
- 相似性推薦：透過針對性的搜尋，像是關鍵字和新聞類別選擇等方式，再透過計算相似性來給予新聞排序，方便使用者快速選取想要觀看的新聞
- 主動性推薦：透過模型的預測，主動性推薦使用者感興趣的新聞。

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，以達到主動性推薦的服務，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，而詳細的分析架構和 code 則放在下方連結當中。<https://github.com/EricTsai83>

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。

Figure 1: 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

Figure 2: News Recommendation System

