

建構拍賣網提問類型之多重標籤辨識模型

Building A Multi-Label Detection Model for Question classification of Auction Website

林怡如 I-Ju, Lin

東吳大學巨量資料管理學院
School of Big Data Management
Soochow University
06770024@gm.edu.tw

吳政隆 Jheng-Long Wu

東吳大學巨量資料管理學院
School of Big Data Management
Soochow University
jlwu@gm.scu.edu.tw

摘要

網路購物已成為現今非常重要的消費型態，每日都會產生成千上萬的銷售問題。此時，客服中心是介於客戶與公司間的第一線服務單位，為了解決客戶服務提問相關問題的需求，本研究運用自然語言處理與機器學習技術，嘗試建構客戶之發問類型辨識模型。其本研究目的為建構一個多重標籤之分類任務資料集，其主是露天採用拍賣之問題類型部分作為主要研究資料來源，以及將資料分為粗細分類的兩組資料集。本研究嘗試以 TF-IDF 和 Word Embeddings 等兩種詞向量演算法進行特徵轉換，以及採用 Extra Trees、Logistic Regression、Random Forest、SVM 等四種機器學習分類演算法進行分類模型建立。本研究實驗結果顯示透 TF-IDF 的詞特徵轉換，以及搭配 Extra Trees 分類模型，其分類效果為最佳，高達 0.8285 F1 分數。顯示露天拍賣網之多重問題類型標籤辨識可有效被機器識別正確。

Abstract

Online shopping (e-commerce) has become an indispensable type of consumption today. The problems faced by the customer service center in e-commerce back-offices are more complicated. This research's core content is to explore how the customer service center to improve the workflow to provide customer's needs solutions by using machine learning

technology. In this study, the detailed and rough types of multi-label detection predicted model was trained by the customers' questions texts about the products of auction website "RUTEN". The TF-IDF and Word Embeddings methods were used to extract the text features, and we experiment with Extra Trees、Logistic Regression、Random Forest、SVM classification models to build a multi-label detection model. The overall result of experiments, the features extracted by the TF-IDF method, and detected by the Extra Trees classification model have performed a better F1 score with 0.82846.

關鍵詞：問題類型分類、多重標籤辨識，文字探勘，詞向量，機器學習。

Keywords: Question Classification, Multi-Label Detection, Text Mining, Word Embeddings and Machine Learning

一、緒論

根據金管會銀行局公開之信用卡消費資料統計，信用卡消費中用來支付網路購物之金額佔比，逐年成長至民國一百零八年來到百分之十八點一，而其EC(Electronic Commerce)消費金額在 108 年來到了新台幣 3,700 億元，與前一年相比成長比達 34%。客戶在網路購物過程中所產生的服務需求問題，無論是網站流量速度（平台系統面）、行銷活動、商品面、物流收送貨、退款速度（金流面）等所有問題，都將全數導向客服中心處理，而客戶需求通常不會是單一存在反而是多樣複雜且難以區分辨識的。在大量進線與留言湧進的同時，單純透過傳統客服人員依循著標準作業流程以及過去的服務經驗值，辨識及處理客戶提出的問題，其處理效率確實非常不佳。在葉靜縈[1]針對壽險業客服對話文本之多標籤主題預測研究中，以客服人員發聲之文本資料的分析結果相對具一定之準確率。由於客服人員具備專業服務訓練，其對話應對被要求必須符合標準作業規範；相較於本研究針對客戶提問內容進行多標籤辨識，口語化文字留言與錯別字的可能性，都增加分析預測的困難度。且因產業屬性的不同，網路購物不論在商品種類、行銷活動、銷售／服務供應鏈等影響因素，都遠比壽險業來得多樣多變。由表 1 客戶提問內容範例可得知，客戶需求通常不會是單一存在反而是多樣複雜且難以區分辨識的。

表 1 客戶提問內容範例

編號	提問內容	類別
1	請問可以貨到付款嗎？可以指定時段嗎？	配送方式、付款方式
2	請問如果要下訂 450 個:1.數量夠嗎？可否挑色？2.希望 7/15 前到貨，是否來的及？	庫存、商品相關、到貨出貨日期

基於上述情況，本研究針對電商客服中心最常見的問與答日常任務，進行自動化辨識客戶需求問題之多重分類模型建置。本研究著採用自然語言處理（Natural Language Processing, NLP）與機器學習技術（Machine Learning）技術，針對電商客服中心收到的客戶需求問題內容，透過詞向量化技術將文本轉換特徵向量，再應用機器學習分類方法建立分類模型，以期望達到對於留言內容進行提問類型的多標籤分類預測，以達成自動分類效果，進而加速回覆客戶的效率與品質為目的。本研究主要貢獻在於產生一組電商平台之提問問題多重標籤分類資料集，能夠讓電商產業業者以此資料集作為雛型進行開發。本研究也嘗試採用不同的文本向量化技術來證實其多重標間分類效果為何，促使學術研究專家學者能夠理解電商平台之多重提問標籤任務的困難度。

二、文獻探討

（一）自然語言處理技術於電子商務問題

客服問題是電子商務領域中相當長常見的議題，多數以建立分類模型等相關研究為主，如在國營事業、金融業、資訊服務業等不同領域的研究中都提到，將客戶提問內容進行建模分類後，客服系統就能依照客戶問題歸屬類別自動分派給權責單位處理，將問題分流指派給專職擅長處理的部門或客服，能快速準確回覆、縮短處理時間；問題被分類系統化進而建立作業標準程序後，使得客服人員能縮短查找該類別問題解答的時間、依循標準作業流程為客戶解答降低錯誤發生，進而分析找出各類別問題產生之原因，並使得後續報表分析能正確地分類。參考過去類似案件的處理經驗回覆眼前客戶的問題，立即且快速、能以好的服務品質提升企業形象以及顧客對商家的滿意度與忠誠度[2-5]。非結構化之文本資料已經成為許多研究議題的主要資料來源之一，文本資料需要轉換為可以被計算的特徵後，才得以應用於各式各樣的求解任務。最常見的特徵轉換方法為 TF-IDF (Term-Frequency Inverse Document Frequency)，是一種統計方法用來評估單詞在許多文章中的重要程度，越重要的單詞越有可能成為關鍵字。而 Word Embeddings 是將文字轉換為詞向量，供後續機器學習分類模型分析使用[6-9]。無論為何種特徵轉換方法，其目的都是為了表達字詞與特定事物的關係。

（二）多分類與多標籤辨識

有關非結構化文本資料多分類預測的相關研究為數不少，多以系統軟體使用者問題、新

聞社群文章、醫療病症、公文文件、客訴問題等單一分類內容居多。相較於本研究處理的是電商客戶提問文本多標籤辨識問題，商品及問題內容方向相對複雜、預測準確困難度較高。對於問題分類的相關研究中，Desai 等人將 Quora 問答平台對語意相似的問題進行分類，其中以 Random Forest 和 XGBoost 等的集成模型提供了最佳性能[10]；Sulaiman 等人則提出有關教育評估的技能和知識，基於認知水平對考試問題進行分類，實驗結果表明 SVC (Support Vector Classification) 有最好的分類效果[11]。而在相關多標籤分類預測研究中，Rokham 等人使用 SVM (Support Vector Machine) 從情緒和精神病性疾病的結構磁共振成像數據中檢測標籤噪聲[12]；Zhou 等人提出了一種多標籤學習方法來聯合學習零件檢測器以捕獲部分遮擋模式，通過共享一組決策樹以增強利用零件相關性[13]。如 PTT 電影版文章[14]與 Google Map POI[15]等多標籤分類的研究，他們嘗試將多標籤分類轉換為單標籤分類問題，分別對各個類別訓練二元分類模型後，測試時將全部分類器的結果組合作為輸出結果，來進行多類別預測分析；相較於本研究以支援 Multi-label Detection 的分類演算法，來預測每一個樣本多個屬性的方式略有不同。它們的研究也將多標籤分類問題進一步針對大小（粗細）類別，分別進行不同難易程度的辨識，其研究結果也是以大（粗）類別的多標籤辨識預測效果較佳。

（三）機器學習分類模型

Logistic Regression 邏輯式迴歸和線性迴歸很類似，都是在確認自變數和因變數之間的關聯。何欣儒[16]在網購低溫食品快速送達與超商取貨服務對購買行為影響之研究中，運用邏輯式迴歸模型，計算分析不同產品類別，使用何種配送方式較符合成本效益。而葉丞峻[17]則以邏輯式迴歸演算法，在解決二元分類不平衡問題有較好的成效。呂育如[18]的實驗在決策變數超過兩種類別時，以邏輯式迴歸模型有較佳的分類效果。隨機森林 (Random Forest) 的名稱是由 Random Decision Forests 而來，隨機森林可以看作一個多棵決策樹的集合，即是一種整合演算法。高詠富[19]在建立財務詐欺檢測模型、陳希聖[20]訓練超音波多特徵脂肪肝疾病分類模型、邱惠君[21]針對社群發言網路霸凌預測，其運用多項機器學習分類演算法進行預測實驗，均以隨機森林演算法能獲得較好的分類效果。Extra Trees 或 Extremely Randomized Trees 演算法由 Geurts 等人[22]提出，此演算法與 Random Forest 十分相似，都是由許多決策樹所組成。與 Random Forest 主要的差異在，Random Forest 是以 Bootstrap 隨機對特徵採樣，作為決策的分裂點，而不是計算最優的相關性的節點進行分裂，之後再基於 entropy 信息增益、gini 基尼係數之類的原

則，選擇一個最佳的特徵值分支屬性。范琪[23]在智能椅活動監控成效評估實驗中，將壓力傳感器採集到的原始數據，使用機器學習分類演算法進行活動分類分析，結果以極限隨機樹的分類精準度高達 98%以上為最佳。在國外相關的研究中，以疾病預測分類於心血管疾病預測系統[24]、在腦腫瘤分類研究中，都使用 **Extremely Randomized Trees** 演算法得到顯著的預測結果，而對於腦腫瘤分類的成效更勝於 **Random Decision Forest** 過去在這個領域的表現[25]。**Support Vector Machine (SVM)** 模型是將訓練資料集中不同類別的各個資料投射為高維特徵空間中的點，透過找到一個超平面將不同屬性類別的點能被盡可能的間隔分開，並使得不同類別之間的邊界間隔距離最大化。接著將測試資料對映到同一個空間中，依照它落在分隔開的哪一側來預測其類別。陳昱瑾[9]針對電商 APP 的使用者／商品／時間特徵來做消費預測，藉以提升 APP 推薦系統的系統效能與推薦效果。實驗結果使用 **RBF** 核函數 **SVM** 預測效果較好且相當，但因 **SVM** 模型預測運算費時，相較之下結構學習法較可以滿足推薦的即時性。**SVM** 分類模型訓練與測試所需時間較長，為了解決這個費時問題，楊鎧謙[15]對 **Google Map POI** 標籤的研究中，嘗試採用 **KDE(Kernel Density Estimation)+SVM** 的混合模型進行多標籤預測，實測時間較單純的 **SVM** 分類模型幾乎快一倍，而預測效果以 **Micro-F1 score** 評估，在大小分類都只略低單純的 **SVM** 模型約 3~7 個百分比。以上分類器是目前對常見的模型，多重標籤分類模型通常採用單一類別模型進行建模和訓練，之後再透過整合每個單一類別模型的機率，就可以達成多重標籤分類任務，因此本研究將嘗試採用多重分類器進行實驗。

三、研究方法

建立提問問題多重標籤分類模型，主要是資料取得與彙整、多重標籤標記、問題內容中文斷詞、問題內容萃取特徵值，以及建模預測四個部分進行說明。其整體概念如圖所示。

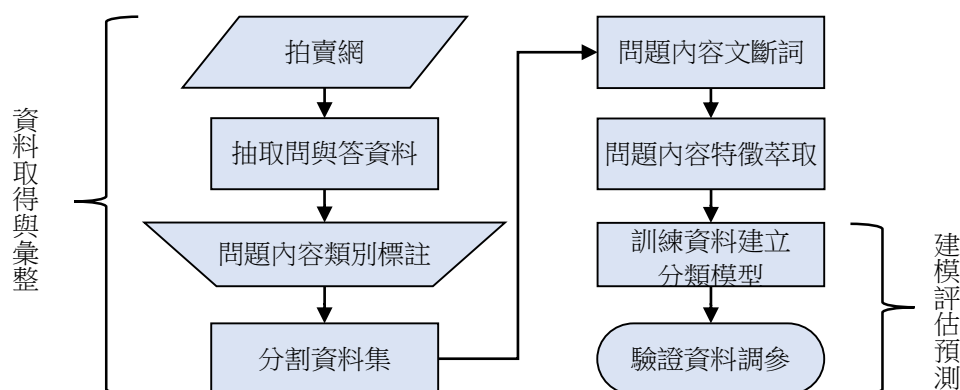


圖 1 本研究方法流程示意圖

(一) 資料取得與彙整

(1) 取得資料與標記

本研究以露天拍賣網客戶對商品的提問內容做為數據來源，主要針對網站首頁分類目錄中，排除成人專區共計 22 類商品類別熱銷商品。針對客戶提問內容進行人工檢閱分類／歸類。本研究參考相關網路購物問題種類，以及實務分析經驗，將客戶對商品提問內容，劃分為十類細分類，並將其歸納為五類粗分類，詳如表 2 所示。逐筆檢閱提問內容後至少須歸屬但不限於一類細分類，再以細分類對應出歸屬之大分類。本研究所設計的粗細分類是為了驗證模型對分類數量的學習效果。

表 2 粗細分類表

五類粗分類	十類細分類	相關項目
金流	發票收據	發票、收據、統一編號、發票抬頭
	價格	確認金額、運費、折扣
	付款方式	付款方式、變更付款方式、退款進度
商品	商品相關	商品規格、使用範圍、保固、包裝、材質、內容物、使用方式、贈品、維修、類似替代品、相關配件、客製化、訂做
	到貨出貨日期	出貨相關、售後查詢配送進度、何時貨到客端
物流	庫存	有沒有貨、何時進貨入庫
	配送方式	面交、取貨方式、自取、變更配送地址、回收
訂單	確認訂單	已下單、確認、修改、取消、退換貨、評價評分、露露通（私訊工具）
	下標相關	下單、加購、合購、預購、平台操作問題、其他賣場、單買
非相關	非相關	問題不完整、非提問

(2) 分割資料集

考量不同商品類別造成客戶提問內容差異化的可能性，為避免各資料集因不同商品類別分配不均影響建模預測結果，本研究將 22 類商品客戶提問資料，依照客戶提問時間先後順序，各自分割為 70%訓練用、10%驗證用、20%測試用，其中 10%的驗證及作為模型調參之使用。

1. 顧客提問之內容中文斷詞

本研究使用 Jieba 演算法對中文資料進行詞語的斷詞。在實驗初期發現，同時建立裝載使用者自訂辭典，以提升斷詞準確效果；並排除標點符號、英文數字字元，以避免因斷詞後字詞數量過多影響分析效率。

2. 問題文本特徵轉換

針對客戶提問內容的特徵值轉換，本研究採用 TF-IDF 與 embeddings 方法，說明如下：

(1) TF-IDF

TF-IDF 是基於頻率的方式將文本轉換，在電商提問的句子，特性用詞有極高機率是特定分類。以訓練資料集斷詞後的字詞作為單一特徵，即計算 TF-IDF 特徵矩陣，除了使用最大詞彙數量，並依據不同詞彙數設定進行多組特徵值實驗效果比較。根據訓練資料集獲得的字詞特徵，分別針對訓練集、驗證集和測試集，轉換為 TF-IDF 特徵矩陣，供後續模型訓練使用。

(2) Word Embeddings

Word embeddings 可以估算其上下文特性，而非只有詞頻，所以在電商提問句子中，常常會出現特性的商品名稱等，所以採用 word embeddings 方法可以有效的辨識出專有名詞的句型。仍是以訓練資料集作為語料庫進行訓練，本計畫分別採用 Word2Vec、FastText、Doc2Vec 詞向量模型進行詞向量訓練。主要的做法是將一整篇的所有字詞的詞向量作加總，因此，每一筆提問的向量就可以產生，即將內容的每個字詞對應的 word embeddings 進行總和計算。

(二) 建模評估預測

在多標籤分類模型部分，本研究運用 Extra Trees、Logistic Regression、Random Forest、SVM 四種機器學習分類演算法進行模型建立。所使用的機器學習演算法將分別對 5 分類與 10 分類任務進行模型訓練。以驗證集進行分類預測後，其分類效果作為最佳參數設定選擇，並以最佳參數設定對測試集進行預測，並評估分類效果。

(三) 評估指標

為了驗證模型在所設計的問題類型辨識效果，以下將針對每個實驗的預測結果，進行評估成效優劣，其評估指標說明如下：

各項評估指標計算方法，統一假設以變數 Q 代表預測筆數、變數 T 代表問題類別數（十類細分類則 $T=10$ 、五類大分類則 $T=5$ ），第 q 筆預測資料的第 t 項問題類別之預測結果計數分別以 $TP_{q,t}$ 、 $FP_{q,t}$ 、 $FN_{q,t}$ 、 $TN_{q,t}$ 表示，其中： TP : True Positive；預測正確，實際提問類別和預測提問類別都是 Yes； FP : False Positive；預測錯誤，實際提問類別是 No、預測提問類別是 Yes； FN : False Negative；預測錯誤，實際提問類別是 Yes、預測提問類別是 No； TN : True Negative；預測正確，實際提問類別和預測提問類別都是 No。

1. 準確率 (Accuracy)

本研究為多重標籤辨識，一筆預測資料中的每一項問題類別必須預測正確為 $TP_{q,t}$ 或 $TN_{q,t}$ 才視為該筆資料預測正確，則第 t 項問題類別的準確率及整體準確率計算如下：

$$Accuracy_t = \frac{\sum_{q=1}^Q TP_{q,t} + TN_{q,t}}{Q}$$

$$Accuracy = \frac{\sum_{q=1}^Q \begin{cases} 1 & \text{if } (\sum_{t=1}^T TP_{q,t} + TN_{q,t}) = T \\ 0 & \text{otherwise} \end{cases}}{Q}$$

2. 精確度 (Precision)

精確度是模型所有的預測值為 Yes 的部分裡，預測正確的效果。但 Precision 並無法反映有多少實際值為 Yes 的樣本被預測錯誤，所以通常不會單獨使用 Precision 指標進行預測效能評估。多重標籤辨識預測第 t 項問題類別的精確度為 $Precision_t = \frac{\sum_{q=1}^Q TP_{q,t}}{\sum_{q=1}^Q TP_{q,t} + FP_{q,t}}$ ，整體精確度計算公式如下。

$$Precision = \frac{\sum_{q=1}^Q \sum_{t=1}^T TP_{q,t}}{\sum_{q=1}^Q \sum_{t=1}^T TP_{q,t} + FP_{q,t}}$$

3. 召回率 (Recall)

又稱 True Positive Rate 或 Sensitivity。召回率是模型所有的實際值為 Yes 的部分裡，被預測正確的程度。和 Precision 相反，Recall 無法反映有多少實際值為 No 的樣本被錯誤預測成 Yes，所以通常也會和其他指標並行評估預測效能。多重標籤辨識預測第 t 項問題類別的召回率為 $Recall_t = \frac{\sum_{q=1}^Q TP_{q,t}}{\sum_{q=1}^Q TP_{q,t} + FN_{q,t}}$ ，整體召回率計算公式如下。

$$Recall = \frac{\sum_{q=1}^Q \sum_{t=1}^T TP_{q,t}}{\sum_{q=1}^Q \sum_{t=1}^T TP_{q,t} + FN_{q,t}}$$

4. F1-Score

F1-Score 是將精準度與召回率合併評估的指標，即精準度與召回率的調和平均值，所以理論上如果 F1-Score 比較高的時候，代表 Precision 和 Recall 也會比較高。本研究以 F1 做為特徵值萃取方法、分類模型、參數設定各種組合後效能評估指標。多重標籤辨識預測第 t 項問題類別的 F1 為 $F1_t = 2 \times \frac{(Precision_t \times Recall_t)}{(Precision_t + Recall_t)}$ ，整體 F1 計算公式如下。

$$F1 = 2 \times \frac{(Micro - Precision \times Micro - Recall)}{(Micro - Precision + Micro - Recall)}$$

四、實驗數據與結果

(一) 實驗資料

本研究針對露天拍賣網 22 類熱銷商品，共爬取 2,108 項商品、84,630 筆發問內容，接續對提問內容進行人工檢閱分類標記，自最新發問內容往回標記，總計共標記 5,760 筆提問資料。人工標記後進行資料集分割，各資料集之問題分類佔比分佈如表 3 和表 4 所示，各資料集客戶提問歸屬類別數分佈如表 5 所示，屬於多標籤問題筆數佔比約 20%。Jieba 斷詞後各資料集字詞數如表 6 所示。

表 3 十分類各資料集之問題分類佔比分佈表

問題資料集	發票收據	商品相關	到貨出貨日期	庫存	非相關	價格	確認訂單	配送方式	下標相關	付款方式	合計
訓練	65	1,474	266	1,457	55	725	198	304	450	137	5,131
驗證	18	202	37	238	-	97	19	43	58	12	724
測試	20	390	81	470	10	186	32	89	112	26	1,416
合計	103	2,066	384	2,165	65	1,008	249	436	620	175	7,271

表 4 五分類各資料集之問題分類佔比分佈表

問題資料集	非相關	物流	金流	商品	訂單	合計
訓練	55	1,889	888	1,474	641	4,947
驗證	-	302	124	202	77	705
測試	10	589	218	390	143	1,350
合計	65	2,780	1,230	2,066	861	7,002

表 5 各資料集客戶提問歸屬類別數分佈表

歸屬類別數	十分類				五分類			
	訓練	驗證	測試	合計	訓練	驗證	測試	合計
1	3,107	454	929	4,490	3,197	465	964	4,626
2	783	101	181	1,065	763	96	171	1,030
3	120	20	31	171	72	16	12	100
4	22	2	8	32	2	-	2	4
5	2	-	-	2				
合計	4,034	577	1,149	5,760	4,034	577	1,149	5,760

表 6 各資料集之字詞數統計分佈表

資料集	字詞總數	平均字數	中間值	眾數	標準差	變異數	標準誤
訓練	4,754	10.52	8	5	9.47	89.60	0.149
驗證	1,206	9.93	8	4	8.28	68.56	0.345
測試	2,029	9.59	7	4	7.81	60.92	0.23

(二) 實驗設計

在實驗設計部分，如表 7 所示，本研究嘗試採用不同大小之 TF-IDF 詞彙數，根據訓練集的資料，最多為 4,754 個單詞，最少使用 50 個單詞。而 Word Embeddings 部分，設計 20 組向量維度設定，以 50 為級距，最大為 1,000。而 4 種機器學習分類演算法之超參數設定值，可以根據表 8 所示。

表 7 詞彙數和詞向量維度設定

向量化方法	設定值
TF-IDF	50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 4754
Word Embedding	50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000

表 8 分類演算法參數設定

分類演算法	參數名稱	參數設定
Extra Trees	決策樹數量 (n_estimators)	100, 200, 300, 400, 500, 750, 1000
	正規化(penalty)	l2
Logistic Regression	求解方式(solver)	lbfgs、newton-cg
	正規強度的倒數(C)	1, 10, 100, 1000, 10000, 100000, 1000000
Random Forest	子樹數量(n_estimators)	100, 200, 300, 400, 500, 750, 1000
SVM	核函數(kernel)	rbf、linear、poly、sigmoid
	scale'(gamma)	0.01, 0.1, 1, 10, auto

※ 其餘未列參數均使用預設值。

(三) 多標籤分類之實驗結果

實驗最後以驗證資料集預測結果進行最佳參數設定選擇，並以最佳參數設定對測試資料集預測結果進行分類預測成效評估。測試集十類標籤分類預測成效評估，各項指標彙整如表 9 所示。整體以 TF-IDF 搭配 Extra Trees 預測效果最佳，F1 分數達 0.82846；Word Embeddings 則以 SVM 演算法有較好的預測效果。

表 9 測試集之十類標籤辨識分成效表

特徵法	分類模型	最佳參數	Accuracy	Precision	Recall	F1
Doc2Vec	Extra Trees	特徵=1000, estimators=400	0.6214	0.9146	0.63560	0.7500
	Logistic Regression	特徵=800, solver=newton-cg, C=1	0.6736	0.8519	0.7797	0.8142
	Random Forest	特徵=750, estimators=400	0.6275	0.9162	0.6483	0.7593
	SVM	特徵=550, kernel=linear, gamma=auto	0.6832	0.8560	0.7811	0.8168
FastText	Extra Trees	特徵=750, estimators=300	0.6171	0.9166	0.6285	0.7457
	Logistic Regression	特徵=600, solver=newton-cg, C=100	0.6658	0.8266	0.7910	0.8084
	Random Forest	特徵=750, estimators=200	0.6240	0.9042	0.64620	0.7537
	SVM	特徵=850, kernel=linear, gamma=auto	0.6710	0.8558	0.7670	0.8089
TF-IDF	Extra Trees	特徵=400, n_estimators=750	0.7172	0.8820	0.7811	0.8285
	Logistic Regression	特徵=3500, solver=newton-cg, C=10	0.6954	0.8720	0.7698	0.8177
	Random Forest	特徵=450, estimators=1000	0.7215	0.8901	0.7719	0.8268
	SVM	特徵=3500, kernel=linear, gamma=auto	0.7102	0.8911	0.7691	0.8256
Word2Vec	Extra Trees	特徵=850, n_estimators=500	0.6310	0.9138	0.6511	0.7604
	Logistic Regression	特徵=650, solver=newton-cg, C=10	0.6641	0.8432	0.7712	0.8056
	Random Forest	特徵=850, estimators=400	0.6414	0.9098	0.6695	0.7714
	SVM	特徵=550, kernel=linear, gamma=auto	0.6728	0.8551	0.7754	0.8133

如表 10 數據所示，測試資料集五類標籤分類預測成效整體仍以 TF-IDF 搭配 Extra Trees 的預測效果最佳，F1 分數達 0.8734；而 Word Embeddings 則以搭配 Logistic Regression 的預測效果較好。

表 10 測試集之五類標籤辨識分成效表

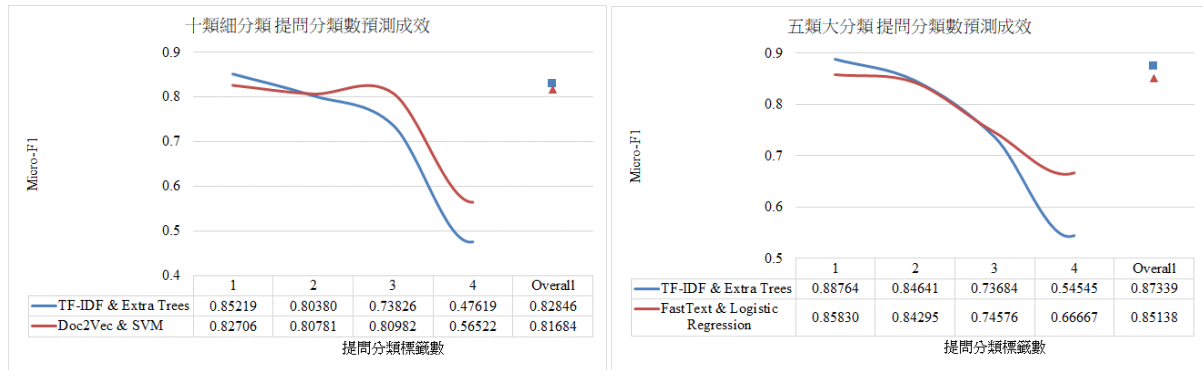
特徵法	分類模型	最佳參數	Accuracy	Precision	Recall	F1
Doc2Vec	Extra Trees	特徵=900, estimators=1000	0.7041	0.9029	0.7437	0.8156
	Logistic Regression	特徵=850, solver=newton-cg, C=1000	0.7285	0.8668	0.8244	0.8451
	Random Forest	特徵=1000, estimators=750	0.7032	0.8987	0.7489	0.8170
	SVM	特徵=750, kernel=linear, gamma=auto	0.7206	0.8728	0.8133	0.8420
FastText	Extra Trees	特徵=900, estimators=1000	0.6928	0.8943	0.7333	0.8059
	Logistic Regression	特徵=850, solver=newton-cg, C=1000	0.7372	0.8584	0.8444	0.8514
	Random Forest	特徵=1000, estimators=750	0.7084	0.8919	0.7578	0.8194
	SVM	特徵=650, kernel=linear, gamma=auto	0.7163	0.8689	0.8148	0.8410
TF-IDF	Extra Trees	特徵=4500, n_estimators=300	0.7807	0.8944	0.8533	0.8734
	Logistic Regression	特徵=4754, solver=newton-cg, C=10	0.7502	0.8867	0.8289	0.8568
	Random Forest	特徵=350, estimators=300	0.7641	0.8764	0.8348	0.8551
	SVM	特徵=4754, kernel=linear, gamma=auto	0.7572	0.8895	0.8289	0.8581
Word2Vec	Extra Trees	特徵=450, n_estimators=100	0.6928	0.9044	0.7356	0.8113
	Logistic Regression	特徵=750, solver=newton-cg, C=10000	0.7102	0.8393	0.8356	0.8374
	Random Forest	特徵=550, estimators=300	0.7041	0.9041	0.7541	0.8223
	SVM	特徵=500, kernel=linear, gamma=auto	0.7111	0.8653	0.8089	0.8361

(四) 結果分析

1. 從客戶提問分類標籤數面向看預測成效

十類細分類、五類大分類分別以 TF-IDF 和 Word Embeddings 演算法預測效果最佳的分類模型預測結果進行比較。如圖 2 所示，單一類別問題被成功預測的效果較高，F1 分

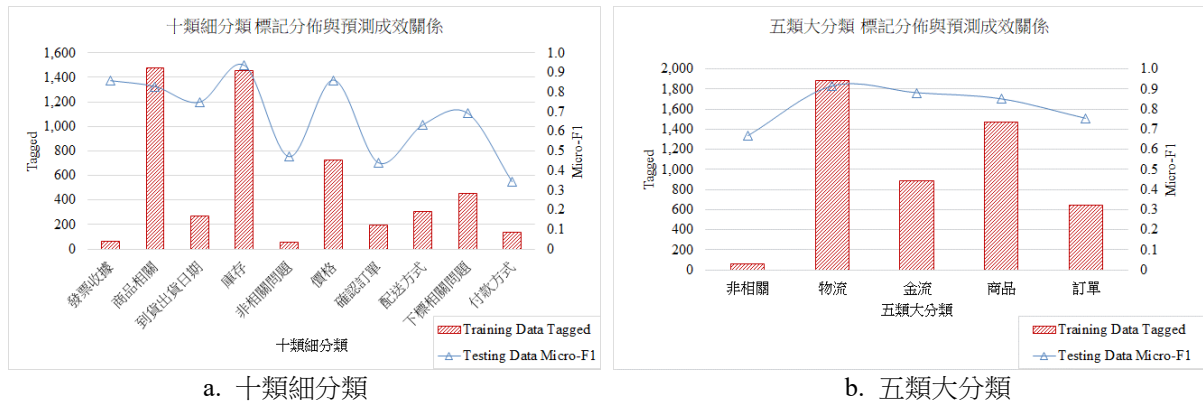
數隨著問題歸屬類別數增加而遞減；而比起 TF-IDF 演算法，對於多重標籤問題辨識，以 Word Embeddings 萃取特徵值的分類模型反而得到了較好的預測效果。



a. 十類細分類不同特徵值萃取法最佳組合比較 b. 五類大分類不同特徵值萃取法最佳組合比較
圖 2 各提問分類標籤數預測成效

2. 從各分類標籤面相看預測成效

以整體預測效果最好的 TF-IDF 搭配 Extra Trees 的預測結果，從分類標籤各自的預測效果來看，對照訓練資料集各分類標籤標記數量，由圖 3 可看出，預測效果較差的分類標籤，對應到訓練資料集中該分類標籤標記數量也相對的較少。



a. 十類細分類 b. 五類大分類
圖 3 標記分佈與預測成效關係

3. 預測結果差異性比較

十類細分類以 TF-IDF、Word Embeddings 特徵值最佳預測效果進行比較。各分類標籤預測優劣比較如表 12 所示，其中 TF-IDF 特徵值預測效果較差的問題類別，Word Embeddings 特徵值反而有較好的預測效果。

表 12 十類細分類各分類標籤預測效果差異比較表

	發票收據	商品相關	到貨出貨日期	庫存	非相關問題	價格	確認訂單	配送方式	下標相關問題	付款方式
TF-IDF	0.8571	0.8269	0.7482	0.9349	0.4706	0.8603	0.4400	0.6323	0.6919	0.3429
Word Embeddings	0.7692	0.8361	0.7733	0.9280	0.5600	0.8362	0.5455	0.6282	0.5864	0.4167

（五）討論

從模型導入實際運用層面來看，首先需要檢視的是：現有分類模型問題類別是否足夠、是否適用，因為每個購物網站的營運模式、作業流程都不盡相同，對於客戶提問分類層級多寡也因其分析需求而有所不同，本研究建議，在模型預測架構實際導入運用時，以本研究提出之十類細分類作為問題類別基礎，依照各平台實務狀況增減調整分類項目，或是擴增另一分類更細微的微分類層級進行運用。從組織層級來區分，十類細分類預測結果，適合提供給客服人員、執行層級，經過分類模型自動預測出客戶問題類型，快速理解客戶需求、縮短人工分派案件時間，進而透過平台系統的整合串接，提供客服人員該問題類別回覆話術參考範本，在客服人員確認調整回覆內容的同時，亦可將正確的提問分類回饋給預測模型，達到校正效果。客服單位透果此模型可以立即發現有疑慮的商品或營運異常狀況，警示相關單位介入處理，有助於在第一時間使得問題獲得解決。

五、結論與建議

本研究整體以 TF-IDF 搭配極限樹得到較好的分類預測效果。對於粗略的五類大分類預測，*F1* 分數可達到 0.8734；而分類較精準、較難辨識準確的十類細分類預測結果也有 0.8285。單一標籤問題預測仍以 TF-IDF 特徵值有較高的 *F1* 分數，而多標籤問題部分，則以 Word Embeddings 特徵值的預測成效較好。由於無法取得現成資料進行實驗，本研究自行撰寫爬蟲程式取得客戶提問文本資料，依實務經驗定義客戶問題分類並人工檢閱標記歸屬類別。本研究嘗試以既有的特徵值萃取演算法及分類模型進行模型建立，實驗結果也顯示分類效果可達八成以上。對此，本研究也創造了電商客戶提問多重標籤辨識資料集，爾後能夠做為相關電商問題之多重標籤研究的參考基礎。

本研究同時使用 TF-IDF 以及 Word Embeddings 兩種特徵值演算法進行實驗比較，雖然整體預測效果以 TF-IDF 較好，但對於多重標籤問題，則以 Word Embeddings 能帶來較高的預測成效，而 TF-IDF 則能對單一標籤案例獲得較佳成效。本研究的成果是在特定拍賣網站下所得分類效果，因此無法直接採用本研究所出的標記原則，但其電商提問多重標籤之標記概念相似，針對特定電商設計特定類別將可以有效建立適合的分類模型。由於缺乏大量電商相關語料庫可供 Word Embeddings 訓練，導致其預測效果無法顯著提升。在未來研究方面，可以採用預訓練的 Word Embeddings 模型，進行詞向量微調，期望有效提升預測效果。例如使用 BERT 的預訓練模型，來強化文本特徵可用性。

參考文獻

- [1] 葉靜縈,「客服對話式文本資料之多主題標籤辨識研究—以某壽險公司為例」,輔仁大學統計資訊學系應用統計碩士在職專班,碩士論文,2019。
- [2] 鄭哲明,「應用資料探勘於顧客問題自動分類之研究-以自來水公司民眾意見信箱為例」,國立交通大學理學院科技與數位學習學程,碩士論文,2015。
- [3] 施瑋蘋,「運用資訊科技輔助客服系統應答最佳化之研究」,國立臺北教育大學資訊科學系碩士班,碩士論文,2019。
- [4] 蔡佩珊,「建置自動化客服回覆機制之聊天機器人」,國立臺北科技大學工業工程與管理系,碩士論文,2019。
- [5] 林益陞,「電子商務網站服務品質、顧客滿意與顧客忠誠之研究 - 以 PChome 為例」,朝陽科技大學應用英語系,碩士論文,2018。
- [6] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, Sentiment analysis based on improved pre-trained word embeddings, *Expert Systems with Applications*, 117, 2019, pages 139-147.
- [7] B. Guo, C. Zhang, J. Liu, and X. Ma, Improving text classification with weighted word embeddings via a multi-channel TextCNN model, *Neurocomputing*, 363, 2019, pages 366-374.
- [8] M. Aydogan, and A. Kaci, Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification, *Physica A: Statistical Mechanics and its Applications*, 541, 2020.
- [9] J. A. Gonzalez, L.-F. Hurtado, and F. Pla, Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter, *Information Processing & Management*, 57, 2020.
- [10] N. Desai, and A. Mahendran, classification of semantically similar question pairs using machine learning. *International Journal of Computing and Digital Systems*, 9, 2020.
- [11] S. Sulaiman, R. A. Wahid, A. Ariffin, and C. Z. Zulkifli, question classification based on cognitive levels using linear SVC. *Test Eng. Manag*, 83, 2020, pages 6463-6470.
- [12] H. Rokham, G. Pearlson, A. Abrol, H. Falakshahi, S. Plis, and V. D. Calhoun, Addressing inaccurate nosology in mental health: A multi label data cleansing approach for detecting label noise from structural magnetic resonance imaging data in mood and psychosis disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2020.
- [13] C. Zhou, and J. Yuan, Multi-label learning of part detectors for heavily occluded pedestrian

- detection. *In Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [14] 黃冠傑, 「多標籤分類方法應用於 PTT 資料」, 淡江大學統計學系應用統計學碩士班, 碩士論文, 2019。
- [15] 楊鎧謙, 「On Large-Scale Multi-Label Classification for POI Tagging」, 國立中央大學資訊工程學系, 碩士論文, 2017。
- [16] 何欣儒, 「網購低溫食品快速送達與超商取貨服務對購買行為影響之研究:二元 logistic Regression 模式之建立與應用」, 銘傳大學企業管理學系, 碩士論文, 2018。
- [17] 葉丞峻, 「適用於分類變數資料的二元不平衡資料自動分類系統」, 淡江大學統計學系碩士班, 碩士論文, 2017。
- [18] 呂育如, 「應用資料探勘分類超音波病人檢查時間」, 中原大學工業與系統工程研究所, 碩士論文, 2019。
- [19] 高詠富, 「法則庫方法萃取財務報表詐欺規則」, 國立雲林科技大學資訊管理系, 碩士論文, 2019。
- [20] 陳希聖, 「使用隨機森林實現超音波多特徵脂肪肝疾病分類」, 國立臺灣大學應用力學研究所, 碩士論文, 2019。
- [21] 邱惠君, 「應用文字探勘技術進行網路霸凌預測」, 國立交通大學管理學院資訊管理學程, 碩士論文, 2019。
- [22] P. Geurts, D. Ernst, and L. Wehenkel, Extremely randomized trees, *Machine Learning*, 63, 3-42, 2006.
- [23] 范琪, 「具有動作辨識之智慧椅研發」, 元智大學通訊工程學系, 碩士論文, 2018。
- [24] R. Shafique, A. Mehmood, S. Ullah, and G. S. Choi, Cardiovascular Disease Prediction System Using Extra Trees Classifier. *Research Square*, 4, 2016.
- [25] M. Goetz, C. Weber, J. Bloecher, B. Stieltjes, H. P. Meinzer, and K. Maier-Hein, Extremely randomized trees based brain tumor segmentation, *Proceeding of BRATS challenge-MICCAI, 006-011*, 2014.