語言模型應用於中文手寫地址辨識
(Language Model Based Chinese Handwriting Address Recognition)

Chieh-Jen Wang, Yung-Ping Tien, Yun-Wei Hung
Service Systems Technology Center, Industrial Technology Research Institute
{chiehjen, JasonTien, joyce_h}@itri.org.tw

摘要

託運單中文手寫地址辨識是智慧物流領域自動化的重要挑戰，而中文手寫字的偵測與辨識是其中的核心。由於手寫字的書寫模式較印刷字複雜多變，辨識上容易誤判，且地址文字在託運單影像中占比小、文字排列緊密，易造成偵測上的困難，因此如何精準偵測託運單上的地址文字是本論文之研究重點。本論文提出託運單地址自動偵測及辨識系統，針對地址字元進行偵測與辨識，透過語言模型降低手寫字誤判的機率，提高辨識正確率。

Abstract

Chinese handwritten address recognition of consignment note is an important challenge of smart logistics automation. Chinese handwritten characters detection and recognition is the key technology for this application. Since the writing mode of handwritten characters is more complex and diverse than printed characters, it is easy misjudgment for recognition. Moreover, the address text occupies a small proportion in the image of the consignment note and arranged closely, which is easy to cause difficulties in detection. Therefore, how to detect the address text on the consignment note accurately is a focus of this paper. The consignment note address automatic detection and recognition system proposed in this paper detects and recognizes address characters, reduces the probability of misjudgment of Chinese handwriting recognition through language model, and improves the accuracy.

關鍵字：手寫辨識、地址辨識、語言模型

Keywords: handwritten recognition, address recognition, language model

1 緒論

新型態商業模式快速發展，加上疫情影響，消費者傾向線上購物，使得物流包裹快速增加，2020 上半年台灣進口包裹數量近五千萬件，為電商帶來 17.5%的銷售額成長。在流通的包裹中，手寫地址託運單佔比仍高，以台灣最大物流處理中心為例，臨櫃交寄託運單手寫地址佔 76%，物流中心每年需處理 500 萬件以上的手寫地址託運單。其託運單種類以紅單(50%)、三聯單(10%)及其他無特定格式之手寫或印刷(40%)為主，因傳統機器無法針對手寫地址進行有效的辨識，因此這些未數位化手寫託運單地址，仍需以大量人工判讀的方式對託運單地址進行分揀，造成人力負擔重，且易產生人為誤判。

即便物流業者開始引入自動分揀、分流系統，但在處理未數位化的手寫地址託運單時，仍需先以人工目視收件地址，將每個託運單的區域代碼輸入分揀機後，才能由機器進行分揀，人力負擔重。為解決物流中心依賴人工判讀託運單地址的問題，本論文提出一套運用自然語言處理及光學影像辨識技術的自動手寫託運單地址偵測及辨識系統，可快速判定託運單寄送區域，提高自動分揀效率。

欲建立自動化的託運單地址辨識系統，關鍵技術在於手寫字的偵測，由於手寫字的書寫模式較印刷字複雜多變，辨識上容易誤判，且地址文字在托運單中占比小、文字排列緊密，造成偵測上的困難，故如何精準偵測託運單上的地址文字是本論文之研發重點。

本研究擬運用自然語言處理技術、語言模型及光學影像辨識技術，針對託運單上的

手寫地址進行辨識，建立託運單地址自動偵測及辨識系統。期待未來可協助物流處理中心達成託運單自動化判定分揀之目的，快速擷取託運單內的寄收件者、地址、電話、公司行號等資訊，加速物流行政處理速度，並減輕物流中心人工判讀託運單地址的負擔，提高託運單分揀效率。

未來除了物流相關產業，自動化的手寫辨識系統還可以擴展到其它不同的領域，例如：可整合 RPA（Robotic Process Automation）(Hofmann et al., 2020)應用，自動擷取訂單資料，利用自動化的手寫辨識技術，辨識不同格式訂單文件，替代人工進行鍵入、複製和貼上等繁瑣且耗時的操作。

2 文獻回顧

過去已經有一些學者對於地址辨識技術做過相關研究，主要是著重在地址語意剖析（Semantic Parser）(Z. Li et al., 2020)，以英文地址為例，通常只要能正確地使用空白斷詞，後續的拼寫校正及格式轉換通常就不會出問題，而不像中文地址常會有資訊錯誤或缺失的情況發生。這些地址剖析相關研究有基於辭典查找(Küçük Matci & Avdan, 2018)、隱藏式馬可夫模型(Hidden Markov Model, HMM) (X. Li et al., 2014)和條件隨機場(Conditional Random Field, CRF)(Arora, 2016; Sun, 2017)，但這幾種模式需要做特徵工程，倚賴專業領域知識去擷取特徵。近期研究(Abid et al., 2018; Sharma et al., 2018)則提出以深度學習為基礎的地址剖析，擺脫上述模型需要特徵工程的限制，也可以提升地址剖析的正確性。

有學者會針對地址文字內容的格式進行正規化(Sun, 2017)，讓所有文字都有統一的表現格式。以地址文字為例包括：阿拉伯數字、數字全形半形及中文數字等各種為了內容及格式統一的正規化。而不同地址型態會套用不同的正規化處理，例如：街路部分的數字要正規化成中文數字，巷弄號則是正規化成半形數字。

語言模型(Chen & Goodman, 1999)經常使用在許多自然語言處理方面的應用，如語音識別，機器翻譯，詞性標註，句法分析，手寫體識別和資訊檢索。人們長期在學習累積語言後有判斷合不合理與聯想的能力，語言模型就是利用機率大小來判斷句子合不合理。換句話說，語言模型描述一個字串 (word sequence, WS) 的機率。N-Gram (Jurafsky et al., 1999)語言模型是基於統計的語言模型算法，主要是將文本中之文字內容，取最靠近的 n 個字當作條件機率計算的先驗條件，形成長度是 n 的字詞片段序列，每個字詞片段及稱為 gram， 常見的 n -Gram 模型有 Unigram(1-gram)，Bigram(2-gram)，Trigram(3-gram)。 但語料庫無法含蓋人類古往今來說過的話，即語料庫的資料量通常不夠，因而許多字串出現的次數為零。神經網絡語言模型 (Neural-network-based language model, NNLM)(Park et al., 2018)。該神經網絡輸入一個字後，能夠輸出字庫中各個字為字串下一個字的機率。不同於 n -gram 統計語言模型從語料庫統計以換算機率，一個神經網絡預測下一個字的機率，是利用訓練的方式得到，如循環神經網絡(Recurrent-neural-network-based, RNN) 語言模型(Xiao & Zhou, 2020)。

3 手寫地址辨識模型

手寫地址辨識模型系統架構如圖 1，可分成 3 個模組，包含：包裹攝影取像、托運單偵測和手寫辨識。利用高速攝影機，直接由輸送帶上拍攝包裹上托運單，再經由影像處理技術(Suri, 2000)，將托運單進行影像特徵擷取(Kumar & Bhatia, 2014)與旋轉校正(Yu et al., 2006)，最後再進行托運單偵測與手寫字辨識。手寫地址辨識系統模型會根據使用者回饋建議修正模型，讓模型辨識正確率隨著時間逐漸提升。

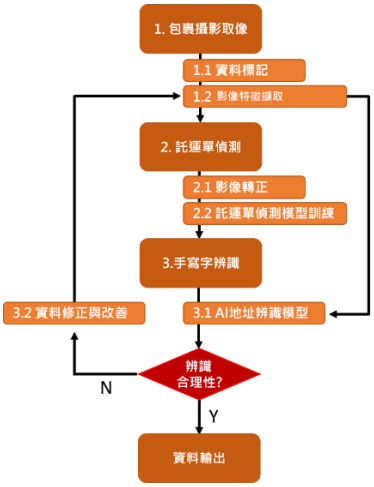

圖 1. 手寫地址辨識模型架構

	Google(美國)	ASTRI(香港)	蒙恬(台灣)	本研究(台灣)
偵測語言	包括英文、中文等 50 種語言	中文	中文、日文、韓文	中文(地址文字)
字元型態	手寫體、印刷體	手寫體	手寫體	手寫體、印刷體
技術	物件偵測、影像分類	影像分類	影像分類	影像分割、物件偵測、影像分類
功能	文字偵測、字元辨識	僅辨識無偵測	僅辨識無偵測	文字偵測、字元辨識
適用情境	適用於書籍或文件影像的解碼	僅適用於字元位置固定的申請表格	行動裝置或電腦的手寫輸入	適用於背景多元的託運單影像中判別地址資訊
優點	支援多種語言	支援手寫中文	支援手寫中文	支援手寫及印刷體的中文地址偵測辨識，正確率高
缺點	對託運單影像中字元的辨識度不高，正確率低	無法進行文字偵測，不適用託運單影像	無法進行文字偵測，不適用託運單影像	特針對中文地址字元進行辨識，通用性略窄

表 1. 全球光學影像辨識系統比較表

3.1 資料收集：託運單拍攝

本研究在台灣最大物流處理中心之物流輸送帶架設高速攝影機，自動拍攝包裹在輸送帶上的托運單，因為輸送帶周圍環境因素(如：輸送帶周圍光源不足與架設像機位子被限制)，相機取像有時會有文字模糊，或因相機取像定位偵測不佳，託運單無法完整拍攝等問題。

本研究總共收集原始影像 15 萬張(如圖 2. 左邊照片)，經過過濾上述問題的照片，再經由人工增加亮度、提高對比度的方式來做調整，有效樣本約 60%，約 9 萬張影像可用。

圖 2. 托運單原始與人工調整影像照片

3.2 託運單偵測

由於輸送帶上包裹並沒有固定的放置位置，也就是說託運單擺放位置並不是固定，因此如何擷取如包裹上面的托運單就是一個重要的議題。本研究透過 YOLO v4(Shafiee et al., 2017)進行託運單偵測， YOLO v4 架構中的 SPP(Spatial pyramid pooling) + PAN(Path Aggregation Network)，可解決小物件偵測問題，有效的偵測出包裹上託運單的位子如圖 3。

圖 3. 託運單偵測

3.3 手寫字辨識

手寫託運單地址辨識一直是相當困難的議題，除不同的人有不同的書寫風格之外，因少掉書寫筆畫順序等重要的特徵來協助辨識，使得離線(offline)手寫辨識相較在線(online)手寫辨識困難許多(Plamondon & Srihari, 2000)。現行文字辨識流程為先進行單一字元偵測如圖 4。再分析字元順序如圖 5。本研究使用 CRNN(Convolutional Recurrent Neural Network)(Shi et al., 2017)進行手寫字辨識；再針對是否有缺漏字等問題，透過地址語言模型進行比對，進行自動填補漏字及合理性驗證，以產出完整地址。

我們使用自然語言處理中語言模型(Chen & Goodman, 1999)，利用地址資料前後文關係建立先驗機率(Prior Probability)模型，自動填補漏字及更正誤判，解決因文字密集排列而造成的字元遺漏問題，語言模型訓練與校正的流程如下。

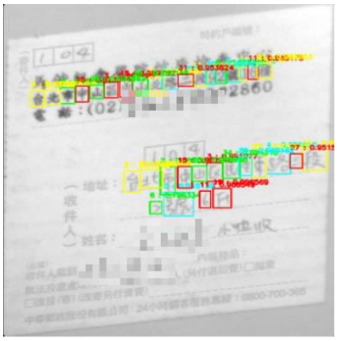

圖 4. 手寫字元偵測

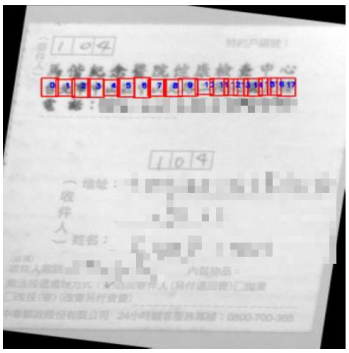

圖 5. 手寫字元排序

- 地址斷詞

地址斷詞採取 CKIP Transformers[1]為基礎架構，再憑藉著過去處理大量地址的經驗，優化成兼具效率及正確性的斷詞模組。輸入一地址資料，模組會產生斷詞地址資訊，如：縣市、鄉鎮市區、村里、路街道、巷弄號等地址斷詞結果。

舉例來說：“新竹縣竹東鎮中興路四段 195 號”，一般可能會斷成<新竹縣 竹東鎮 中興路 四段 195 號>，因為把‘中興路’當成斷詞的關鍵字，而我們則能成功斷成<新竹縣 竹東鎮 中興路四段 195 號>。

- 地址正規化

為了訓練語言模型，所以先將所有地址資料庫內的文字進行正規化，包括：阿拉伯數字、數字全形半形及中文數字等各種內容及格式統一的正規化。而不同地址型態會套用不同的正規化處理，例如：街路部分的數字要正規化成中文數字，巷弄號則是正規化成半形數字。

- 語言模型訓練

因為手寫字不同於印刷體文字，辨識出的地址通常都是有缺漏或是帶有些許錯誤，所以還是需要透過語言模型技術，比對輸入地址與校正基準語言模型中的地址，找一個最接近且唯一地址做為輸出的地址。本研究使用自己收集的地址資料外，也使用戶政司公開資料來訓練語言模型，利用 *n-gram* (Jurafsky et al., 1999)語言模型的編碼技術去達成高速的字串的校正比對，篩出候選地址後，再進行編輯距離（edit distance）的計算來找到最接近的候選地址。

4 效能評估

本研究主要有三個階段需要進行評估，如圖 6：包含托運單偵測、手寫字元偵測和手寫字元辨識。由地址資料庫中隨機挑選 6,000 張地址影像進行測試，文字影像標記字元數約達 10 萬字。

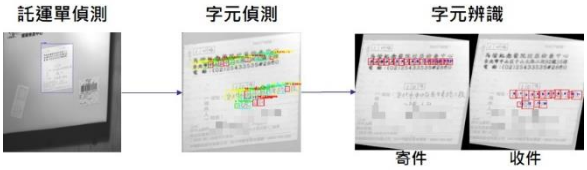

圖 6.三階段評估

我們使用正確率為評估指標，評估指標如公式如下：

$$\text{正確率} = \frac{\text{模型正確辨識字元數}}{\text{影像中的地址字元數}}$$

訓練過程正確率與損失率曲線如圖 7，可以觀察到隨著 epoch 增加，正確率跟著提升而損失率隨著下降。

在校能比較方面，我們使用目前業界辨識效能最好的 Google Cloud Vision API[2]當比較基準(Baseline)，系統效能如表 2，可以發現在托運單手寫字辨識應用，本研究所提出模型效能優於 Google Cloud Vision API，主要的原因可能是因為有針對手寫字與包裹托運單資料進行優化。有使用語言模型進行地址文字校正可以將正確率由 70%提升到 84%，可以發現使用語言模型進行地址校正對於系統效能是非常有幫助。

[1] https://github.com/ckiplab/ckip-transformers

[2] https://cloud.google.com/vision

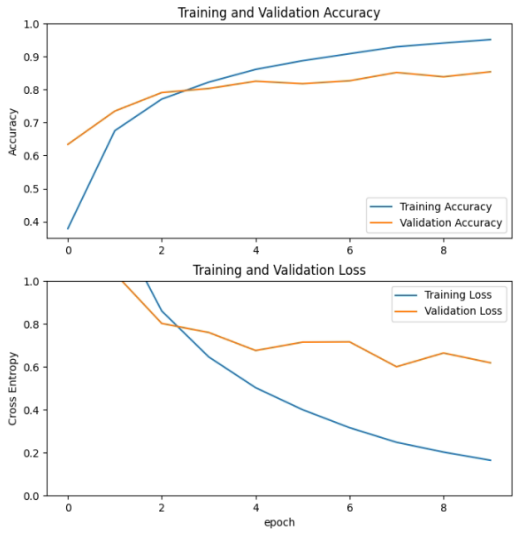

圖 7. 訓練過程正確率與損失率曲線如圖

	Google API	本研究	本研究 (語言模型)
託運單偵測	---	82.74%	82.74%
字元偵測	70%	83.06%	83.06%
字元辨識	41.17%	70%	84%

表 2. 托運單手寫字辨識效能

對於辨識錯誤的案例如圖 8，系統會將「台」辨識為「6」或將「號」辨識為「路」，我們進行錯誤分析，發現因為手寫字撰寫較為擁擠、字跡草亂、有塗抹等情況，易造成文字辨識上的誤判。後續改善將以整行文字的模式進行偵測與擷取，先將以斷行的模式進行文字擷取後，再進行分析。用來改善字元定位並計算文字順序的問題，可降低文字定位失誤而產生的文字缺失問題。也可以嚐試調整物件偵測及文字辨識模型架構，建立相關模型再訓練機制，針對手寫字再進行資料蒐集與訓練。

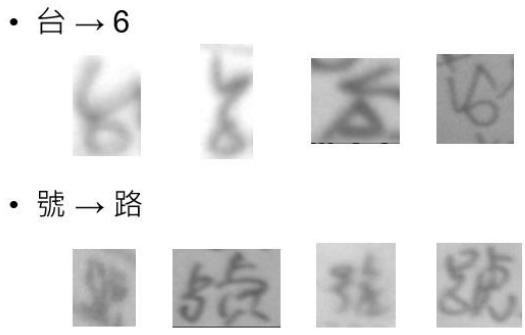

圖 8. 系統辨識錯誤實例

5 結論與未來規劃

現行包裹手寫地址佔比仍高，物流中心每年需處理 500 萬件以上的手寫地址包裹，因傳統機器無法針對手寫地址進行有效的辨識，因此這些未數位化手寫包裹地址，仍需以大量人工判讀的方式對包裹地址進行分揀，造成人力負擔重，且易產生誤判。

本研究建立自動化的包裹地址辨識系統，關鍵技術在於手寫字的偵測與辨識，由於手寫字的書寫模式較印刷字複雜多變，辨識上容易誤判，且地址文字在托運單中占比小、文字排列緊密，造成偵測上的困難，故如何精準偵測包裹上的地址文字是研發重點。

輸入包裹影像，手寫辨識系統先偵測託運單再進行字元偵測，根據字元偵測結果，計算影像傾斜角度並進行影像轉正，使用字元座標進行文字排序後，進行字元辨識並輔以語言模型根據地址關鍵字(縣、市、區等等)進行漏字判斷，自動填補漏字，最後輸出托運單地址。經過 6,000 筆測試資料驗證，模型效能優於 Google Cloud Vision API。

本研究運用人工智慧技術、小物件偵測及密集物件偵測技術，針對包裹託運單上的手寫地址進行辨識，建立包裹地址的自動偵測及辨識系統，期待未來可協助物流處理中心達成包裹自動化判定分揀之目的，並減輕物流中心人工判讀包裹地址的負擔，提高包裹分揀效率。

參考文獻

Abid, N., ul Hasan, A., & Shafait, F. (2018). DeepParse: A Trainable Postal Address Parser. *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 1–8. https://doi.org/10.1109/DICTA.2018.8615844

Arora, N. (2016). Knock knock: Who's there? package delivery at the right address. *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*, 86–89. https://doi.org/10.1145/3007818.3007828

Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, *13*(4), 359–394. https://doi.org/10.1006/csla.1999.0128

Hofmann, P., Samp, C., & Urbach, N. (2020). Robotic process automation. *Electronic Markets*,

30(1), 99–106. https://doi.org/10.1007/s12525-019-00365-8

Jurafsky, D., Martin, J. H., Kehler, A., Linden, K. V., & Ward, N. (1999). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*.

Küçük Matci, D., & Avdan, U. (2018). Address standardization using the natural language process for improving geocoding results. *Computers, Environment and Urban Systems*, *70*, 1–8. https://doi.org/10.1016/j.compenvurbsys.2018.01.009

Kumar, G., & Bhatia, P. K. (2014). A Detailed Review of Feature Extraction in Image Processing Systems. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 5–12. https://doi.org/10.1109/ACCT.2014.74

Li, X., Kardes, H., Wang, X., & Sun, A. (2014). HMM-based Address Parsing with Massive Synthetic Training Data Generation. *Proceedings of the 4th International Workshop on Location and the Web*, 33–36. https://doi.org/10.1145/2663713.2664430

Li, Z., Qu, L., & Haffari, G. (2020). *Context Dependent Semantic Parsing: A Survey* (arXiv:2011.00797). arXiv. https://doi.org/10.48550/arXiv.2011.00797

Park, S., Song, J.-H., & Kim, Y. (2018). A Neural Language Model for Multi-Dimensional Textual Data based on CNN-LSTM Network. *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 212–217. https://doi.org/10.1109/SNPD.2018.8441130

Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 63–84. https://doi.org/10.1109/34.824821

Shafiee, M. J., Chywl, B., Li, F., & Wong, A. (2017). *Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video* (arXiv:1709.05943). arXiv. https://doi.org/10.48550/arXiv.1709.05943

Sharma, S., Ratti, R., Arora, I., Solanki, A., & Bhatt, G. (2018). Automated Parsing of Geographical Addresses: A Multilayer Feedforward Neural Network Based Approach. *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, 123–130. https://doi.org/10.1109/ICSC.2018.00026

Shi, B., Bai, X., & Yao, C. (2017). An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2298–2304. https://doi.org/10.1109/TPAMI.2016.2646371

Sun, W. (2017). Chinese named entity recognition using modified conditional random field on postal address. *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–6. https://doi.org/10.1109/CISP-BMEI.2017.8302311

Suri, J. S. (2000). Computer Vision, Pattern Recognition and Image Processing in Left Ventricle Segmentation: The Last 50 Years. *Pattern Analysis & Applications*, *3*(3), 209–242. https://doi.org/10.1007/s100440070008

Xiao, J., & Zhou, Z. (2020). Research Progress of RNN Language Model. *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 1285–1288. https://doi.org/10.1109/ICAICA50127.2020.9182390

Yu, Z., Dong, J., Wei, Z., & Shen, J. (2006). A Fast Image Rotation Algorithm for Optical Character Recognition of Chinese Documents. *2006 International Conference on Communications, Circuits and Systems*, *1*, 485–489. https://doi.org/10.1109/ICCCAS.2006.284682