

應用記憶增強條件隨機場域與之深度學習及  
自動化詞彙特徵於中文命名實體辨識之研究

**Leveraging Memory Enhanced Conditional  
Random Fields with Gated CNN and Automatic  
BAPS Features for Chinese Named Entity Recognition**

簡國峻\*、張嘉惠\*

**Kuo-Chun Chien and Chia-Hui Chang**

**摘要**

命名實體辨識是在自然語言處理當中一個重要的任務。現今基礎深度學習模型應用於資料品質較為優良的命名實體擷取，雖有不錯的效果，但在社群媒體資料集中卻未能達到傳統條件隨機場域之基準值。由於一個命名實體有能可多次在文中提及，因此藉由上下文資訊來改進命名實體的擷取也是近年來的研究方向。在本研究中，我們延伸記憶增強條件隨機場域 **MECRF** 於中文的命名實體擷取，利用門控卷積網路及雙向 **GRU** 網路來增強記憶條件隨機場域，以利模型抓取長距離的文章資訊。此外，也藉由特徵探勘擷取命名實體前後詞彙以及前綴後綴詞彙特徵（簡稱為 **BAPS**），並使用模型可自動訓練的參數，自動調整詞向量及 **BAPS** 詞彙特徵。最後我們同時採用字元及詞彙向量來增進模型的效能。本研究所提出之模型，在網路社群媒體的人名辨識資料中可以達到的 91.67% 準確率，在 **SIGHAN-MSRA** 中也得到最高的 92.45% 地名實體辨識效果及 90.95% 整體召回率。

**關鍵詞：**機器學習、命名實體辨識、神經網路、特徵探勘

---

\* 國立中央大學資訊工程學系

Department of Computer Science & Information Engineering, National Central University

E-mail: qk0614@gmail.com; chia@csie.ncu.edu.tw

## Abstract

Named Entity Recognition (NER) is an essential task in Natural Language Processing. Memory Enhanced CRF (MECRF) integrates external memory to extend Conditional Random Field (CRF) to capture long-range dependencies with attention mechanism. However, the performance of pure MECRF for Chinese NER is not good. In this paper, we enhance MECRF with Stacked CNNs and gated mechanism to capture better word and sentence representation for Chinese NER. Meanwhile, we combine both character and word information to improve the performance. We further improve the performance by importing common before and common after vocabularies of named entities as well as entity prefix and suffix via feature mining. The BAPS features are then combined with character embedding features to automatically adjust the weight. The model proposed in this research achieve 91.67% tagging accuracy on the online social media data for Chinese person name recognition, and reach the highest F1-score 92.45% for location name recognition and 90.95% overall recall rate in SIGHAN-MSRA dataset.

**Keyword:** Machine Learning, Named Entity Recognition, Memory Network, Feature Mining

## 1. 緒論 (Introduction)

命名實體辨識(Named Entity Recognition, NER)是自然語言處理中訊息理解的第一步，其目標是提取當中的命名實體並歸類到預先定義的分類當中，如：人名、地名、組織等。傳統的機器學習於命名實體的辨識任務中，大多使用統計式條件隨機場域進行序列標記，因此受限於小範圍的特徵擷取。如何在中文的資料集當中擷取參考長距離上下文資訊，判斷當前字詞正確的語意，進而正確的辨識命名實體，是機器理解訊息根本的任務。

近年來深度學習被運用在序列標記的模型建立，得到不錯的進展。例如 Huang 在序列標記的任務上使用長短期記憶(Huang, Xu & Yu, 2015)，應用於英文的資料集當中獲得了非常好的效能。Liu 等人於 IJCNLP 2017 將記憶網路的概念加入條件隨機場域當中(Liu, Baldwin & Cohn, 2017)，提出 MECRF 架構，透過整合上下文額外的記憶，使模型能夠獲取較長範圍以外的文章特徵，同樣在英文資料集上獲得了出色的表現。然而這些基礎深度學習模型應用於資料品質較為優良的資料集上，雖均有不錯的效果，但在社群媒體資料集中卻未能達到傳統機器學習方式之基準值，因此如何有效地擷取文字中所隱含的資訊，使模型有較好的濾除雜訊之能力，也是在應用上非常重要的一環。

為改善上述的限制，本研究延伸記憶增強條件隨機場域 MECRF 於中文命名實體辨識任務；MECRF 的概念是基於上下文可能不只一次提及實體名稱，以及 Attention 機制的應用，藉以更正確找出命名實體。我們首先透過訓練詞向量模型，將字元轉換為數值

化之資料；再藉由卷積層、雙向 GRU 層提供模型更多的特徵，及整合長距離文章資訊的記憶層，使命名實體任務不同於往常僅能夠擷取小範圍的資訊，能夠獲取豐富完整的文章訊息。此外，也藉由特徵的探勘(Chou & Chang, 2017)，並使用深度學習模型可自動訓練的參數，自動調整詞向量及詞彙特徵，除長距離的文章資訊外，更能充分獲得文章所隱藏的訊息。

本研究所使用的資料為 Chou 及 Chang 所使用的 PerNews 測試資料集，但其資料集是以句子為單位進行標記，並無上下文，因此我們自製爬蟲程式，蒐集原始資料的網路新聞及社群媒體做為訓練及測試資料。經實驗結果比較，在網路社群媒體的資料中可以達到的 91.67% 的標記準確率，與尚未加入記憶的模型相比大幅提升 2.9%，再加入詞向量及詞彙特徵，與基礎的記憶模型相比更是提升了 6.04%。本研究提出之模型也在 SIGHAN-MSRA 中得到最高的 92.45% 地名實體辨識效果及 90.95% 召回率。

## 2. 相關文獻回顧 (Related Work)

序列標記已經發展許久，常見的模型有隱藏式馬可夫模型(Hidden Markov Model, HMM)、最大化熵馬可夫模型(Maximum Entropy Markov Model, MEMM)以及條件隨機場域(Conditional Random Field, CRF)。Lafferty 等人(2001)所提出的條件隨機場域在自然語言處理序列標記(Sequence Labeling)的任務中，是多數人的選擇且被廣泛的應用，但是條件隨機場域僅能夠抓取小範圍的文章資訊(Finkel, Grenager & Manning, 2005)，對於獲取整篇文章中的資訊則是條件隨機場域關鍵的限制。

### 2.1 卷積神經網路(Convolutional Neural Networks)

卷積神經網路是一種前饋神經網路，通常由卷積層(Convolutional)、池化層(Pooling)、全連接層(Fully-Connected)組成，相較於其他的網路，卷積神經網路所需要使用的參數較少，因而成為一種頗具吸引力的深度學習模型。卷積神經網路擁有能夠自動抓取相鄰特徵的優點，Collobert 等人(2011)首先將卷積神經網路自動抓取相鄰特徵的優點應用在自然語言處理的序列標記任務中，讓自然語言處理不再相依於專業知識特製而成的特徵模板。近期，Wang 等人(2017)透過堆疊式的卷積神經網路更有結構、多階層地萃取中文語意特徵，同時結合 Dauphin 等人(2016)提出的閘門線性單元(Gated linear unit, GLU)，應用於中文斷詞任務中。

### 2.2 遞歸神經網路(Recurrent Neural Networks)

RNN 則是另一種處理序列型輸入的神經架構，但是單純的 RNN 模型無法擷取長距離的文章資訊，為了不受局部限制的影響，因此有常短期記憶(Long Short Term Memory)的提出。Huang 等人(2015)在序列標記的任務上使用長短期記憶，導入雙向(Bidirectional)的概念來擷取正向及反向的資訊，應用於英文的資料集當中獲得了非常好的效能。

但是遞歸神經網路隨著輸入句子的長度增加(Cho, van Merriënboer, Bahdanau &

Bengio, 2014)，會帶來效能的惡化。在相關的研究(Lai, Xu, Liu & Zhao, 2015; Linzen, Dupoux & Goldberg, 2016)更顯示，遞歸神經網路包括其變化之類型，儘管已經加入時間序列的標記，但仍偏向於相鄰的字元資訊，在涉及遠程上下文依賴性的判斷中表現不佳。

### 2.3 記憶網路(Memory Networks)

傳統的條件隨機場域沒有能力去抓取較長範圍以外的文章特徵，而遞歸神經網路在長距離的文章資訊擷取上效能也並不出色，因此，Weston 等人提出記憶網路(Memory Network)來增強擷取長範圍文章特徵的表現，並應用於問答(QA)的任務當中(Weston, Chopra & Bordes, 2014)，證明記憶的增加對於執行需要常距離文章資訊的推理至關重要。

近期，Liu 將記憶網路的概念加入條件隨機場域當中(Liu *et al.*, 2017)，透過整合額外的記憶(Memory)，使模型能夠獲取較長範圍以外的文章特徵，並且在英文資料集上獲得了出色的表現。

## 3. 模型架構及方法(Model Architecture and Method)

在命名實體辨識標記任務中，每一個句子  $S$  是由  $T$  字元(character)組合而成的序列  $S = \{w_1, \dots, w_T\}$ ，其對應的標籤序列可表示為  $Y = \{y_1, \dots, y_T\}$ 。不同於傳統的條件隨機場域僅需要輸入句子，MECRF 的特點是另有上下文資訊或稱之為記憶體  $M_s$ 。假設每篇文章是由  $|D|$  句子組成  $D = \{S_1, \dots, S_{|D|}\}$ ，與其對應的序列標籤集合  $L = \{Y_1, \dots, Y_{|D|}\}$ 。為避免輸入整篇文章造成記憶體消耗過大，我們僅抓取當前句子  $S_t$  的前後  $B$  句共抓取  $2B+1$  個句子  $M_s = \{S_{t-B}, \dots, S_t, \dots, S_{t+B}\}$  做為短期記憶(short context)，其長度可記為  $N = \sum_{i=t-B}^{t+B} T_i$  (其中  $T_i$  表句子  $S_i$  的長度)。每個輸入字元  $w_j$  可以透過 word2vec 或 GloVe 對文字進行編碼，以  $EMB(w_j)$  來表示。假設  $D$  為 Embedding 的維度，則短期記憶增強隨機場域的輸入序列為大小  $T \times D$  的句子  $E^S$ 、和  $L \times D$  的短期記憶  $E^M$ 。

### 3.1 Stacked CNNs with Gated Mechanism

在本篇論文中，我們應用多層次卷積 (Convolution Layer) 來萃取文字特徵，並參考 Dauphin 等人做法在層與層間加入門控機制來泛化萃取的特徵。門控機制廣泛地應用於循環神經網路架構中，用來控制長期神經網路中資訊的流動；在卷積神經網路中雖沒有長期依賴的問題，不需要輸入閥門以及遺忘閥門，但是 Dauphin 等人認為在多層次的卷積神經網路中，層與層之間可以透過類似輸出閥門的門控機制來決定神經元的流通與否，並有效率地擷取有效的特徵。假設前面嵌入層輸出為  $E^S$  ( $\epsilon R^{T \times D}$ )，則此處卷積運算可表示為：

$$A = E^S \oplus W_K + b \quad (1)$$

其中  $W_K$  為大小為  $K \times D$  的卷積運算過濾器 Kernel Filter， $K$  若過小導致不能含括有效資訊；若過大導致含括冗餘資訊對系統產生不必要的干擾，本研究中將  $K$  設定為 3，再透過多層卷積層擴及字元前後資訊；我們將滑動視窗移動的格數(strides)設為 1，並將補零方式

(padding)設為 SAME，意即輸出長度等於輸入長度。此處卷積運算後不採用池化層 (Pooling Layer)，其原因為中文語意中每個特徵都有其意義，不像影像可能會經過放大、縮小或者位移，因此本研究直接將 L 個卷積 Filters 輸出的 feature maps 做連接 (Concatenation)。

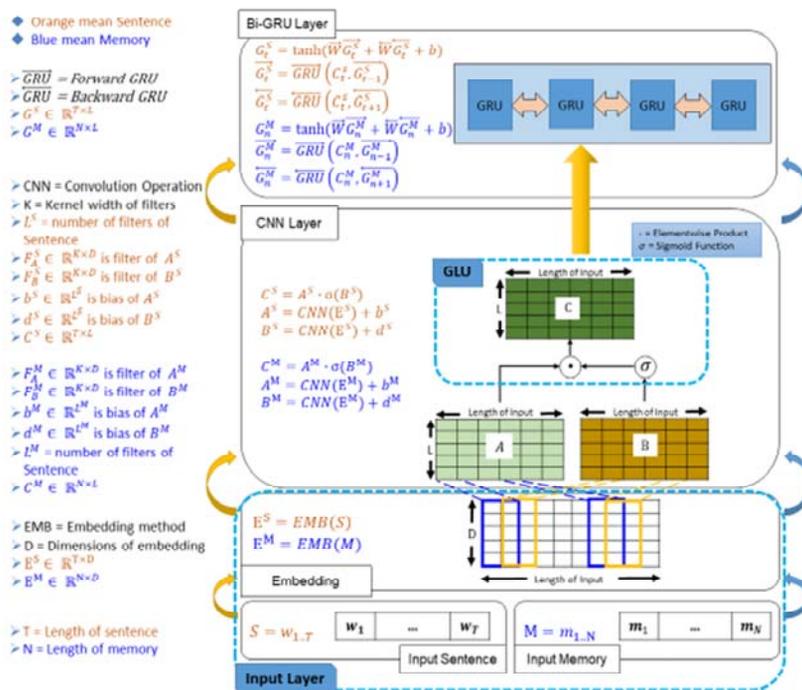


圖 1. Input sentence and memory representation

令 A 及 B 分別為經由兩組 CNN 卷積運算之後所產生的矩陣。前者不經過任何啟動函數，後者將通過一非線性轉換(sigmoid function)用來決定神經元的取捨，再將兩輸出做矩陣逐元素乘法(element-wise multiplication)，如式 (1)所示。

$$C = A \cdot \sigma(B) \quad (2)$$

應用多層卷積 (Stacked Convolution) 及門控機制 (Gated-CNNs) 擷取相鄰字詞特徵後，我們參考 MECRF 採用遞歸神經網路(RNN)的變化體 GRU，且透過雙向的技術來擷取當下位置處的文字 $C_t$ 正向及反向的資訊，並且於輸出時，將正向及反向的資訊套用一個非線性單元 tanh，做為位置 t 的輸出資訊 $G_t$ ，如式(2)。

$$G_t = \tanh(\vec{W}\vec{G}_t + \overleftarrow{W}\overleftarrow{G}_t + b) \quad (3)$$

如圖 1 所示，我們以 $C^S$ 及 $C^M$ 分別代表句子 S 及記憶 M 經過兩組卷積層後的輸出， $G^S$ 及 $G^M$ 分別代表句子 S 及記憶 M 經過雙向 GRU 層後的輸出。

### 3.2 記憶層(Memory Layer)

我們參考 MECRF 做法，使用二組雙向長短期記憶(LSTM)分別對記憶 GM 進行編碼，將時間序列訊號加入模型當中，產生輸入記憶(Input Memory)以及輸出記憶(Output Memory)，如式(4)、(5)。

$$I_j = \tanh(\overrightarrow{LSTM}(G_j^M) + \overleftarrow{LSTM}(G_j^M)) \quad (4)$$

$$O_j = \tanh(\overrightarrow{LSTM}(G_j^M) + \overleftarrow{LSTM}(G_j^M)) \quad (5)$$

假設當前輸入是句子的第  $t$  個字元  $G_t^S$ ，為了計算  $G_t^S$  與記憶當中每個元素  $I_j$  的注意力值  $A_{t,j}$ ，我們將當前輸入  $G_t^S$  與輸入記憶  $I_j$  做內積運算，但是不同於 MECRF 採用 Softmax，此處我們採用  $\tanh$  函數強化重要的記憶位置，如式(6)，其中  $j \in [1, N]$ 。

$$A_{t,j} = \tanh((G_t^S)^\top I_j) \quad (6)$$

最後使用加權和來計算當前的輸出  $p_t$ ，並結合當前輸入  $G_t^S$ ，做為最後的輸出，如式(8)。

$$p_t = \sum_{j=1}^N A_{t,j} O_j \quad (7)$$

$$U_t = G_t^S + p_t \quad (8)$$

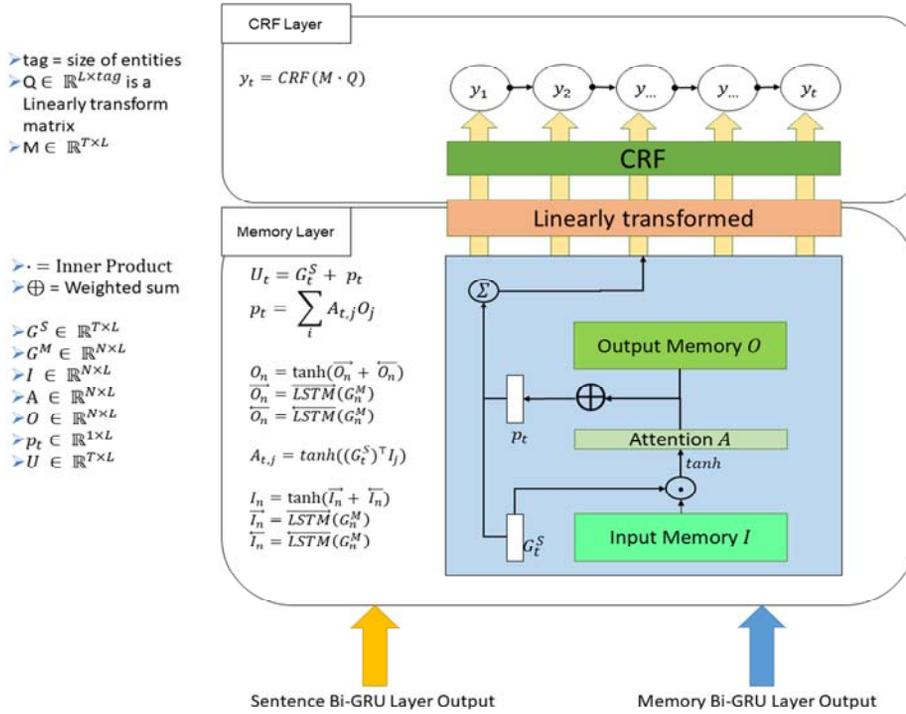


圖2. Memory Enhanced Model with CRF output layer

Attention 的機制允許模型可以不受限制的訪問文章中短期記憶涵蓋的位置，讓我們的模

型可以獲取較豐富的文章資訊。最後我們採用條件隨機場域，經由轉移矩陣考慮標記之間的依賴關係，用以增加準確率。完整架構可參考圖 2。

#### 4. 實驗與系統效能(Experiments and System Performance)

在本章節中，我們將針對本研究所提出的各層模組效能及可調整的變數進行比較。我們使用 PerNews 及 SIGHAN-MSRA 兩組資料評估模型之效能。其中 PerNews 資料集(Chou & Chang, 2017)係以辨識社群媒體上的人名實體為主要研究目標，藉由 7053 個人名清單，自動標記出現於資料集中的人名。不同於 Chou 與 Chang 的做法僅留包含人名的句子，本研究因需要參考上下文資訊，因此在標記時不會過濾掉未含有任何實體的句子。SIGHAN-MSRA 則是學術界普遍用來評估中文斷詞與命名實體辨識工具效能的標準數據集(Levow, 2006)，本研究主要針對人名、地名、組織名進行命名實體辨識。資料集之基本統計資料如 Table 1 所示。

**Table 1. PerNews and SIGHAN-MSRA Datasets**

Dataset	Sentences	Average Characters/ per Sentence	Person	Location	Organization
PerNews Train	335,056	13.13	54,338	N/A	N/A
PerNews Test	363,572	13.14	54,546	N/A	N/A
SIGHAN-MSRA Train	141,546	14.94	17,615	36,861	20,584
SIGHAN-MSRA Test	11,679	14.45	1,973	2,886	1,331

本研究採用的標記法為 BIESO 標記法，評估方式為精準比對完整命名實體後，以常用的指標，即精確率、召回率以及 F1-Score 來進行效能的評估。模型所採用的參數如 Table 2 所示：中文字元嵌入維度 250、三層卷積層、每層 50 個 Kernel Filters、短期記憶體為 200 字元，學習率與 dropout rate 分別為 0.0005 及 0.2。

**Table 2. Model Hyper-Parameters**

Hyper-parameters	value
Character Embedding	250
Conv layer # filters	50
Kernel width of filters	3
Learning rate	0.0005
Dropout rate	0.2
Memory size	200

## 4.1 PerNews Dataset

首先我們針對本篇提出的門控多層卷積雙向 GRU 資訊表示方式，與 DS4NER 工具(Chou & Chang, 2017)所提供的基於前後字詞及首尾字詞特徵的 CRF++方法進行比較。如 Table 3 所示，DS4NER 搭配 CRF++工具的效能僅有 0.8603，而單純使用字元嵌入的 CE-ME-CRF 效能也僅有 0.8572 左右，顯示並非只採用記憶架構就能達到好的字詞及句子的表達方式有其重要性。我們發現多層卷積雙向 GRU 架構優於單純 CE-ME-CRF 效能 2.1%，加入短期記憶的模型更有效提升 2.9%F1。

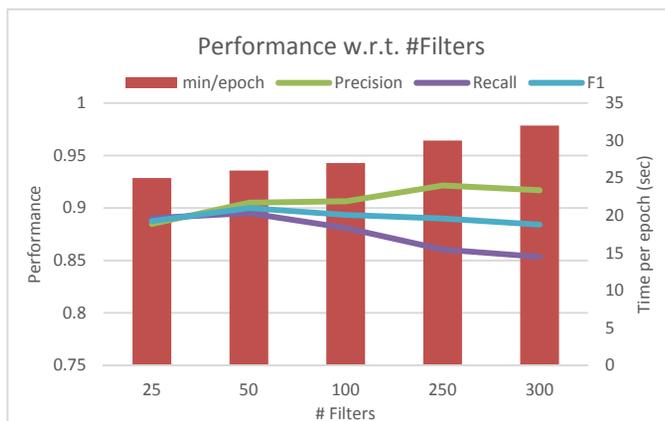
**Table 3. Performance on PerNews Dataset**

Models	Min/epoch	Precision	Recall	F1
DS4NER-CRF++	25*	0.9347	0.7968	0.8603
CE-ME-CRF	15	0.8881	0.8284	0.8572
CE-CNNs-BIGRU-CRF	25	0.9345	0.8289	0.8785
CE-CNNs-BIGRU-ME-CRF	65	0.9067	0.9084	0.9075

\* DS4NER-CRF++為全部訓練時間

### 4.1.1 卷積層過濾器數量 (Number of Filters of Convolution Layer)

在此實驗中，我們調整卷積(CNN)層的過濾器的數量，比較不同過濾器數量對於效能的影響。如圖 3 所示，將過濾器數量設定為 50 的時候，效能表現最佳，過濾器數量逐漸增加的情形下，並無法顯著提升效能，且值得注意的是，越多的卷積過濾器雖因產生更多的特徵，可以得到較好的精準率，但是召回率的表現上則是逐步下滑。



**圖 3. 不同過濾器數量對於效能的影響**  
 [Figure 3. Effects of Filters Number for CNN Layer]

### 4.1.2 卷積層數(Number of Convolution Layers)

在此實驗中，我們調整卷積(CNN)層數量，比較不同卷積層數量對於效能的影響。如圖 4 所示，卷積層數量增加時，對於命名實體的召回率可以有效的提升，當層數設定為 3 的時候，效能表現最佳。

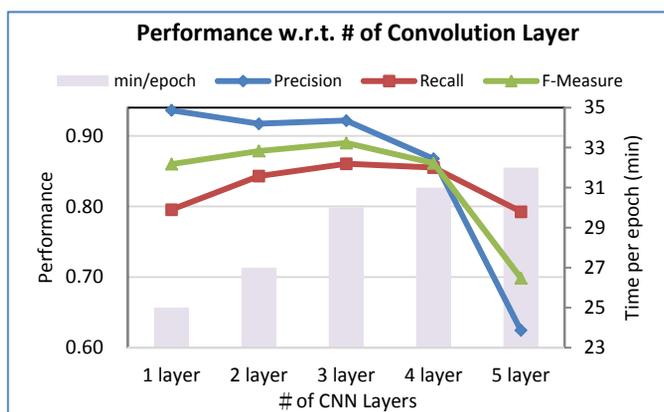


圖 4. 卷積層數對於效能的影響  
[Figure 4. Effects of CNN Layers]

### 4.1.3 記憶體大小 (Memory Preparation Method and Memory Size)

在這個實驗中，我們不僅比較記憶體大小對效能的影響，同時也比較從文章起始做為參考上下文對於效能的影響。如圖 5 所示，採用前後 B 句相較從文章起始的短期記憶方法，效能較佳。我們以 3 句、 7 句，分別對應 100、200 字元以及全文 300 字元進行實驗。由於 PerNews 為網路上之資料，當中擁有許多雜訊，因此在記憶過大的情況下，模型參考到較多的雜訊資料，效能反而有所減損，而在 200 字元記憶體時效能最佳。

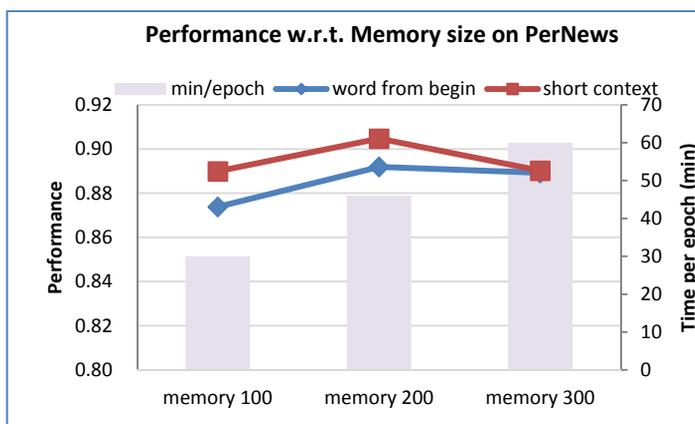
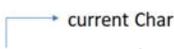


圖 5. 記憶體與效能的影響  
[Figure 4. Effects of Memory Size]

#### 4.1.4 加入詞向量 (Word Embedding)

雖然堆疊卷積可以找出相鄰字元之間的關係，但是其實無法像中文詞彙詞向量那麼有意義，因此我們在基於字元的標記當中加入以當前字 $w_i$ 為中心，與前後字元結合，加入適當的詞向量。如式(8)所示，我們主動加入 $w_{i-1}w_i$ 、 $w_iw_{i+1}$ 、 $w_{i-2}w_{i-1}w_i$ 、 $w_{i-1}w_iw_{i+1}$ 、 $w_iw_{i+1}w_{i+2}$ 等五個詞的詞向量，若查無此詞彙，則補零向量。

$$X_i = [V_{char}(w_i), V_{word}(w_{i-1}w_i), V_{word}(w_iw_{i+1}), V_{word}(w_{i-2}w_{i-1}w_i), V_{word}(w_{i-1}w_iw_{i+1}), V_{word}(w_iw_{i+1}w_{i+2})] \quad (9)$$


  
 外交部昨天證實  
 (The Ministry of Foreign Affairs confirmed yesterday)

Position	Char	Kind	Combination	Words
Pre_2	外	Bichar_Pre	Pre_1 + Cur	交部
Pre_1	交	Bichar_Suf	Cur + Suf_1	部昨
Cur	部	Trichar_Pre	Pre_2 + Pre_1 + Cur	外交部
Suf_1	昨	Trichar_Mid	Pre_1 + Cur + Suf_1	交部昨
Suf_2	天	Trichar_Suf	Cur + Suf_1 + Suf_2	部昨天

**圖 6. 詞彙詞向量產生方法**  
*[Figure 5. Illustration of Adding Word Embedding]*

如圖 6，以「部」為範例，在新增的五個詞彙向量，僅有「外交部」這個詞有對應向量，其於四個字詞均以零向量取代。

在詞向量前處理中，我們採用結巴斷詞系統的精確斷詞模式，再使用 CBOW 建立 Word2vec 模型(Mikolov, Chen, Corrado & Dean 2013)，設定詞頻為至少出現 5 次，訓練 50 維的詞彙(word)的詞向量模型。

#### 4.1.5 自動前後字詞典特徵(Automatic BAPS Dictionary-Based Features)

由於 PerNews 在原始資料中，是藉由 Chou and Chang (2017)所提出的方式進行特徵的探勘，找出 Common Before、Common After、Entity Prefix、Entity Suffix 等特徵(Support 閾值設定為 0.5)，做為 Dictionary-based Features (簡稱 BAPS)，得到不錯的效能。因此我們試圖將此四類特徵各三種長度(1-gram, 2-gram, 3-gram)共 12 個特徵，經過 CNN-BiGRU-MECRF 與上述模型結合，再使用一個可由模型自動訓練的變數 $\alpha$  ( $\alpha \in [0,1]$ )來調整嵌入向量(EMB)與 BAPS 特徵所佔的比重，經過式(10)的計算後，最後再使用條件隨機場域進行序列標記。

$$output = \alpha \cdot EMB + (1 - \alpha) \cdot BAPS \quad (10)$$

換言之，BAPS 特徵也同樣經過多層卷積雙向 GRU 以及記憶網路的計算，學習得新的特徵（如圖 7 所示），最後與字元向量與前面加入的字元及詞向量表示（EMB）統整為條件隨機場域的輸入。

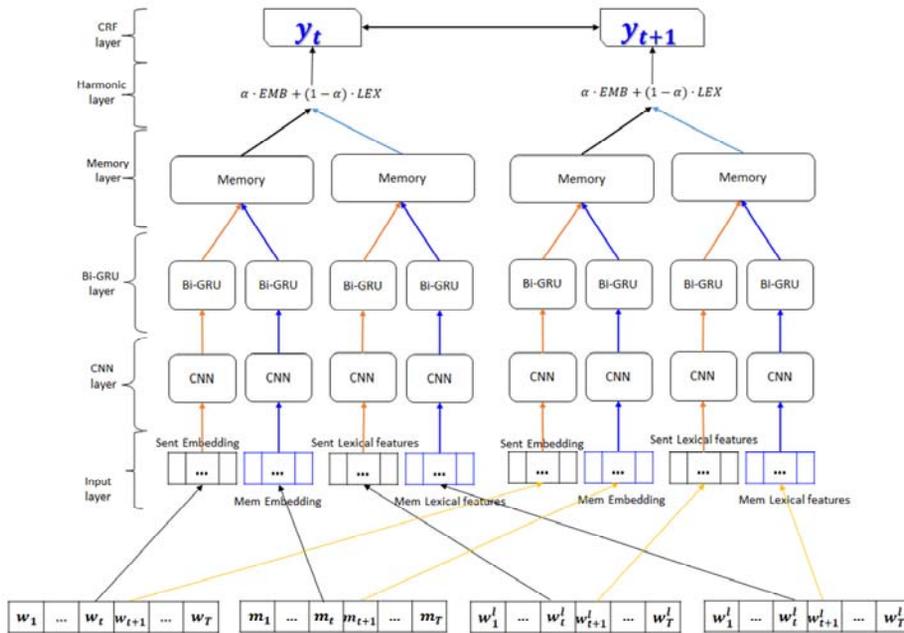


圖 7. Embedding and BAPS Hybrid Model

加入詞向量與 BAPS 特徵的效能如 Table 4 所示，兩者各別改進的幅度不大，綜合來看有 1% 的進步，讓 F1 效能達到 0.9176。

Table 4. Effects of adding word embedding and dictionary-based features

Models	Min/epoch	Precision	Recall	F1
CE-CNNs-BIGRU-ME-CRF	65	0.9067	0.9084	0.9075
CWE-CNNs-BIGRU-ME-CRF	101	0.9467	0.8779	0.9110
BAPS-CNNs-BIGRU-ME-CRF	31	0.9134	0.7951	0.8502
Harmonic (CWE+BAPS) -CNNs-BIGRU-ME-CRF	137	0.9307	0.9048	0.9176

## 4.2 SIGHAN-MSRA 資料集

在 SIGHAN-MSRA 資料集上，實驗結果如 Table 5 所示。本研究所提出的調和(Harmonic)模型，引入詞向量以及自動前後字詞典特徵的資訊，在效能上均有不錯的表現。此外，在地名的評估效能高達 92.45%，在整體召回率也達到了 90.95% 的出色表現。

Table 5. Performance on SIGHAN-MSRA

Model	PER-F	LOC-F	ORG-F	Precision	Recall	F1
Zhou-CRF (2006)	90.09	85.45	83.10	88.94	84.20	86.51
Chen-CRF (2006)	82.57	90.53	81.96	91.22	81.71	86.20
Zhou (2013)	90.69	91.90	86.19	91.86	88.75	90.28
Zhang-MEMM (2006)	<b>96.04</b>	90.34	85.90	<b>92.20</b>	90.18	<b>91.18</b>
Dong-BiLSTM-CRF (2016)	91.77	92.10	<b>87.30</b>	91.28	90.62	90.95
Liu-MECRF (2017)	91.09	91.87	83.81	89.16	90.47	89.81
Lex-CNNs-BiGRU-MECRF	81.70	75.00	67.22	85.30	67.33	75.26
CWE-CNNs-BiGRU-MECRF	91.92	90.84	84.76	89.44	90.16	89.80
Harmonic(CWE+BAPS)-CNNs-BiGRU-MECRF	92.70	<b>92.45</b>	86.31	91.34	<b>90.95</b>	91.14

## 5. 結論與未來展望(Conclusion and Future Work)

本研究所提出的模型，除了使用門控式多層卷積層來自動編碼鄰近字詞外，再使用 Bi-GRU 增加對上下文序列的資訊擷取功能，達到較佳的語意表示，最後使用記憶增強來加強長距離的文意擷取效能，充分獲得文章中所隱含的資訊，可找出有效的特徵做為序列標記的判斷依據。相較於 Liu 等人所提的基本 MECRF 記憶模型而言，本研究所提出的模型在社群媒體資料集中更具有穩定性及效能，而將模型應用於資料品質較好的官方資料上，同樣也有優良的效能展現。

加入 BAPS 特徵探勘所獲得的資訊雖然可能增加部份效能，是否對不同語言有同樣的效能，是後續我們想要探討的地方；另外文字在句子中的位置能否做為一種資訊，或許可以透過特別的編碼來達成。最後，由於近年來深度學習於語言領域應用日廣，對於文字的理解能否有通用的解法，也是未來努力的方向。

## 參考文獻(References)

- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In *Proceedings of Eighth*

- Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, 103-111.
- Chou, C.-L. & Chang, C.-H. (2017). Mining features for web ner model construction based on distant learning. In *2017 International Conference on Asian Language Processing (IALP)*, 322-325. doi: 10.1109/IALP.2017.8300608
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.*, 12, 2493-2537.
- Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2016). Language Modeling with Gated Convolutional Networks. CoRR abs/1612.08083
- Dong, C., Zhang, J., Zong, C., Hattori, M., & Di, H. (2016). Character-Based LSTM-CRF with Radical-Level Features for Chinese Named Entity Recognition. In *Natural Language Understanding and Intelligent Applications - 5th CCF Conference on Natural Language Processing and Chinese Computing (NLPPCC 2016), and 24th International Conference on Computer Processing of Oriental Languages (ICCPOL 2016)*, 239-250.
- Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating Nonlocal Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL '05)*, 363-370. doi: 10.3115/1219840.1219885
- Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF Models for Sequence Tagging. CoRR abs/1508.01991
- Lafferty, J., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01)*, 282-289.
- Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent Convolutional Neural Networks for Text Classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*, 2267-2273.
- Levow, G.-A. (2006). The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In *Proceedings the Fifth SIGHAN Workshop on Chinese Language Processing*, 108-117.
- Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *TACL*, 4 (2016), 521-535. doi: 10.1162/tac1\_a\_00115
- Liu, F., Baldwin, T., & Cohn, T. (2017). Capturing Long-range Contextual Dependencies with Memory-enhanced Conditional Random Fields. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP 2017)*, 555-565.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. CoRR abs/1301.3781
- Sun, J. (2012). "Jieba" (Chinese for "to stutter") Chinese text segmentation: built to be the best Python Chinese word segmentation module.

- Wang, C., Chen, W., & Xu, B. (2017) Named Entity Recognition with Gated Convolutional Neural Networks. In Proceedings of Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, (NLP-NABD 2017, CCL 2017), 110-121. doi: 10.1007/978-3-319-69005-6\_10
- Wang, C. & Xu, B. (2017). Convolutional Neural Network with Word Embeddings for Chinese Word Segmentation. CoRR abs/1711.04411
- Weston, J., Chopra, S., & Bordes, A. (2014). Memory Networks. CoRR abs/1410.3916