

基於參數生成網路的遷移學習進行情感分析和歌手命名識別

Aspect-Based Sentiment Analysis and Singer Name Entity Recognition using Parameter Generation Network Based Transfer Learning

曾筱雯 Hsiao-Wen Tseng 張嘉惠 Chia-Hui Chang 莊秀敏 Hsiu-Min Chuang
中央大學資訊工程學系 中央大學資訊工程學系 國防大學理工學院資訊工程學系
wen80187@g.ncu.edu.tw chia@csie.ncu.edu.tw showmin1205@gmail.com

摘要

當訓練資料有限時，如何應用已標記的訓練資料，幫助目標任務的模型快速建構，是遷移式學習 (Transfer Learning) 的重要議題。在本論文中，以多任務學習 (Multi-task Learning) 的方式進行中文歌手命名實體辨識 (Name Entity Recognition, NER) 和基於面向的情感分析 (Aspect-Based Sentiment Analysis, ABSA) 的任務。我們應用參數生成網路 (Jia et al., 2019) 結合梯度反轉層 (Gradient Adversarial Layer, GRL) (Ganin and Lempitsky, 2015) 架構來建立模型，並且使用 Tie/Break 規則進行標記，動態調節權重的機制 (Dynamic Weight Average, DWA) (Liu et al., 2019)，依據每個任務的損失變化率來調整任務權重。實驗結果顯示，我們的擴展參數生成網路模型 (Extended Parameter Generation Network, E-PGN)，在僅考慮 NER 任務時，F1 可以達到 90%，和 IBHB 效能 86% 相比，有所改善，加入 ABSA 任務後，平均 F1 能夠達到 78%，和 IBHB (Chiu, 2020) 效能相差了 22%，明顯的大幅成長。

Abstract

When we are interested in a certain domain, we can collect and analyze data from the Internet. The newly collected data is not labeled, so the use of labeled data is hoped to be helpful to the new data. We perform name entity recognition (NER) and aspect-based sentiment analysis (ABSA) in multi-task learning, and combine parameter generation network (Jia et al., 2019) and DANN architecture (Ganin and Lempitsky, 2015) to build the model. In the NER task, the data is labeled with Tie, Break, and the task weight is adjusted according to the loss change rate of each task using Dynamic Weight Average (DWA) (Liu et al., 2019). This study used two different source domain data sets. The experimental results show that Tie, Break can improve the results of the model;

DWA can have better performance in the results; the combination of parameter generation network and gradient reversal layer can be used for every good learning in different domain.

關鍵字：參數生成網路、梯度反轉層、命名實體辨識、目標情緒分析、多任務學習。

Keywords: Parameter Generation Network, Gradient Adversarial Layer, Named Entity Recognition and Aspect-Based Sentiment Analysis

1 簡介

自從網路進入 Web2.0 時代，各種部落格及社交平台的興起，人們主動地將他們的想法、意見發布在網路上，讓我們更容易在網路上得知目前討論的熱門話題、流行趨勢。這些使用者原創內容 (User Generated Content, UGC) 變成與情分析的消息來源。當我們對某個主題有興趣時，便可以蒐集這些資料，並透過分析文本提到的目標，以及對目標的情緒來進行進一步的分析。

命名實體辨識 (Named Entity Recognition, NER) 是資訊抽取 (Information Extraction, IE) 技術的任務之一，目標是識別文本中的實體 (Entities)。然而，社群網路上的文章由於是廣大使用者所分享的想法與意見，所以這類的資料與報章雜誌相較下，文句較無正規的文法，且經常包含有表情符號、火星文、注音文等非正規寫法，在用字遣詞上，常有用戶自創的詞彙。本論文所使用的目標任務資料集是從批踢踢實業坊 (Ptt.cc) 所爬取音樂相關的評論，因此在文本分析上更具挑戰性。

情感分析 (Sentiment Analysis) 則是判別文句的情感極性，表示作者對意見目標的正向、反向或中立的立場。本研究主要結合命名實體辨識與情感分析，即意見目標擷取與情感分析兩項任務。本論文的情感分析問題屬於基於面向的情感分析 (Aspect-Based Sentiment Analysis, ABSA) 任務，ABSA 旨在辨別文句中面向 (Aspect) 的情感極性。這個議題對於不少

企業來說相當重要。由於更為細粒度 (Fine-grained) 分析其產品在不同面向的網路評價，更有助於產品的持續改善。

然而在標記資料有限時，如何應用現有其他領域已標記的訓練資料幫助加速目標任務的模型建構，是一項重要的研究議題，也是遷移式學習 (Transfer learning) 主要解決的問題。換言之，遷移式學習是借用來源域 (Source Domain) 中的知識遷移到目標域 (Target Domain)，提高模型在目標域的表現，這樣便可以減少大量目標域訓練資料的依賴，同時提升模型在目標域的表現。

在本論文中，我們使用遷移式學習來改善歌手名稱辨識及情感分析模型的效能，結合參數生成網路 (Parameter Generation Network, PGN)(Jia et al., 2019) 和梯度反轉層 (Gradient Adversarial Layer, GRL)(Ganin and Lempitsky, 2015) 的模型，分別針對語言模型 (Language Model, LM)、命名實體辨識和面向的情感分析三個任務進行多任務學習 (Multi-task Learning)。由於多任務學習在計算整體損失 (Loss) 時，有多個子任務損失，每個任務的資料數量大小不一，因此在參數的設定上，我們參考了動態計算權重 (Dynamic Weight Average, DWA)(Liu et al., 2019) 的方式，以達到最好的效能。在資料集的標記準備上，為了能有效分隔中文詞彙，有別於常見的 BIEOS 標記，我們採用 Tie/Break 的標記方式 (Shang et al., 2018)，表達中文字與字之間是否連結在一起，再預測每個連結的詞是否為命名實體。

實驗結果顯示，我們提出的擴展參數生成網路模型 (Extended Parameter Generation Network, E-PGN)，在使用新聞人名資料做為來源域時 (Chou et al., 2016)，在僅考慮歌手 NER 任務和語言模型任務時，則 NER 效能可以達到 90% 的 F1，相比 IBHB 效能 86% 改進了 4.65%；若是加入 ABSA 任務一起考慮時，則在歌手 NER 任務上達到 85.5%，相較 IBHB 效能 87.4% 雖略有下降，但在 ABSA 任務可以達到 78% 的 F1 效能，相較 IBHB(Chiu, 2020) 效能 56% 相差 22% 是相當大的突破，即使新聞人名資料並不包含 ABSA 標記資訊。

2 相關研究

遷移式學習強調來源域和目標域中的領域 (Domain) 和任務 (Task) 的概念。Pan and Yang (2010) 定義了領域和任務。領域 $D = \{X, P(X)\}$ 是由特徵空間 X 和特徵空間上的邊際機率分布 $P(X)$ 組成。 $X =$

$\{x_1, x_2, \dots, x_n\} \in X$ ， X 是一個文本資料， x_i 則是文本中對應的第 i 個字的向量。一般來說，如果兩個域不同，那他們大多是具有不同的特徵空間或是不同的邊際機率分布。任務 $T = \{Y, P(y|x)\}$ 由標籤空間 Y 和目標預測函數 $P(y|x)$ ， $Y = \{y_1, y_2, \dots, y_n\} \in Y$ ， Y 是一個文本資料 X 的標籤， y_i 則是文本中對應的第 i 個字的標籤。透過來源域 D_S 和來源域任務 T_S 中的知識來改進目標域 D_T 中目標預測函數 $P(y|x)$ 的學習，其中 $D_S \neq D_T$ ，或 $T_S \neq T_T$ 。

在 ABSA 任務上，使用不同領域的資料來建立模型的像是 Hu et al. (2019) 專注於提取領域不變特徵，He et al. (2019) 引入了消息傳遞架構，並使用不同領域的文本級資料建立模型，Li et al. (2019) 將選擇性對抗學習作為領域適應方法引入到模型中，Zhang et al. (2019) 基於 LSTM 的交互式注意力轉移網路，實現跨領域的 ABSA。

2.1 參數生成網路 (Parameter Generation Network, PGN)

Jia et al. (2019) 在進行跨域的命名實體辨識實驗中，設計了一種新的參數生成網路 (PGN)，使用了領域嵌入 (Domain Embedding) 的方式，對不同領域進行處理。透過領域和任務之間的相似性，來學習領域和任務相關性和領域及任務嵌入 (Task Embedding)。

參數生成器 主要的模型是使用了 BiLSTM-CRF，透過領域嵌入和任務嵌入生成 BiLSTM 參數 $\theta_{LSTM}^{d,t}$ ，以達到在不同領域、不同任務上轉移知識的目的。

$$\theta_{LSTM}^{d,t} = W \otimes I_d^D \otimes I_t^T \quad (1)$$

式 1 中， d 表示來源域和目標域， t 表示命名實體辨識和語言模型兩個任務， $I_d^D \in R^U$ 是領域嵌入， U 是領域嵌入大小， $I_t^T \in R^V$ 是任務嵌入， V 是任務嵌入的大小， W 是由 $P(LSTM)*V*U$ 組成的三維張量， $P(LSTM)$ 則是 LSTM 的參數數量，使用 \otimes 來進行張量縮減 (Tensor contraction)。

2.2 神經網路的域對抗訓練

Ganin and Lempitsky (2015) 提出 DANN 模型，DANN 是受生成對抗網路 (Generative Adversarial Network, GAN) 所啟發的對抗技術，生成和訓練資料集分佈一致的資料。使用對抗層辨別特徵的來源，當對抗網路的表現不佳時，表示來源域和目標域的特徵只存在細微差異，可遷移性更好，反之亦然。

DANN 的模型架構，由以下三個部分組成：

特徵提取器 將來源域和目標域的資料都映射到特徵空間上，混合兩個域的資料後，並提取後續網路需要完成任務所需要的特徵。

標籤分類器 標籤分類器的資料為來源域帶有標籤的資料，並對特徵空間來源域的資料進行分類，分辨出正確的標記，協助特徵提取器所提取的特徵是能夠用來做分類。

域分類器 對特徵空間的資料進行域分類，分辨出數據是來自哪個領域，是由不帶標籤的來源域和目標域的資料組成。

域分類器和標籤分類器的輸入都是來自特徵提取器，不過域分類器的目標是最大化域分類損失，混淆來源域和目標域的資料；標籤分類器的目標則是最小化分類損失，讓標籤可以精準的被分類。

為了解決兩個分類器的目標不同，Ganin and Lempitsky (2015) 提出了梯度反轉層，在反向傳遞的過程中，梯度方向自動相反，在前向傳遞過程中恆等變換。也就是在域分類器損失的梯度反向傳遞到特徵提取器的參數之前，自動加一個負號，如此便實現了和生成對抗網路相似的對抗損失。

3 模型架構

本論文的模型架構主要是基於 PGN (Jia et al., 2019) 模型架構進行擴展，因此本研究的模型命名為 Extended Parameter Generation Network，簡稱為 E-PGN。

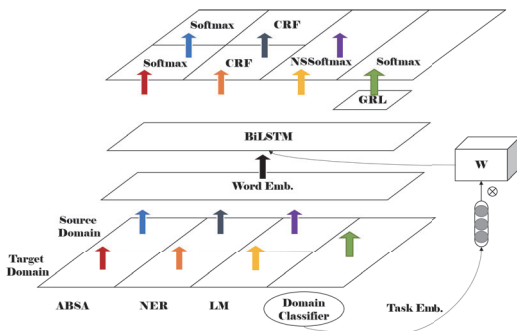


Figure 1: E-PGN 模型架構圖

模型架構如圖 1，在原 PGN 模型基礎上，保留任務參數生成的架構，且除了命名實體辨識及語言模型任務外，增加了 ABSA 任務；另外在領域參數生成的部份，則不使用 Domain Embedding，而改以領域對抗式 (Domain Adversarial) 架構，結合梯度反轉層，增加一個混淆領域分類的損失函數。我們在後面小節中分別詳述本模型的各層功能。

3.1 輸入層

輸入層共有七個資料集，每個資料集在輸入資料時，皆是以句子的字元 (Character-based) 為單位。NER 和 ABSA 任務的來源域表示為 $S_{NER/ABSA} = \{x_i, y_i\}_{i=1}^m$ 和目標域 $T_{NER/ABSA} = \{x_i, y_i\}_{i=1}^n$ ，LM 任務的來源域和目標域分別表示為 $S_{LM} = \{x_i\}_{i=1}^m$ ， $T_{LM} = \{x_i\}_{i=1}^n$ ，領域分類任務表示為 $D_{Dom} = \{x_i, y_i\}_{i=1}^{m+n}$ 。其中 x 代表句子內的第 i 個字， y 代表句子內第 i 個標記， m, n 表示為資料集的第 m, n 個句子。

3.2 嵌入層

從輸入層得到的字元，在嵌入層轉換為數值向量，本論文使用 GloVe (Global Vectors for Word Representation)(Pennington et al., 2014)，將 GloVe 模型預先訓練的詞向量表示為 V_c ，並使用查表的方式映射，找到該字元所表示的向量。映射函數寫為

$$v_i = V_c(X_i) \quad (2)$$

其中 X_i 是輸入的字元在 V_c 中的第 i 個， v_i 則表示通過 GloVe 所得到的詞向量且 $v_i \in R^D$ ， D 為向量維度。

3.3 參數產生層

參數產生器可以讓不同領域間知識能夠傳遞，動態地生成 θ_{LSTM}^t 參數，不同於原 PGN 的設計，在 θ_{LSTM}^t 部分，我們只保留了任務嵌入。

$$\theta_{LSTM}^t = W \otimes I_t \quad (3)$$

式 3 中， t 表示為任務，可以為語言模型、命名實體辨識、ABSA、域分類四種任務。 $I_t \in V$ 表示為任務嵌入， V 是任務嵌入的大小， $W \in P(LSTM) * V$ 是由 BiLSTM 的參數數量和任務嵌入的大小組成。最後 W 和 I 使用 \otimes 計算得到 θ ， \otimes 用來進行張量的縮減 (Tensor Contraction)。

3.4 雙向長短記憶網路層

雙向長短記憶網路 (Bi-directional Long Short-Term Memory, BiLSTM) 由前向 LSTM 與後向 LSTM 組合而成。為了最大化考量文章中上下文的資訊，本論文採用 BiLSTM 的方式來進行實驗，使用雙向的訊息傳遞 $\vec{h}_i^{d,t}$ 和 $\overleftarrow{h}_i^{d,t}$ ，表示如式 4。

$$\vec{h}_i^{d,t} = LSTM(\vec{h}_{i-1}^{d,t}, v_i, \vec{\theta}_{LSTM}^{d,t}) \quad (4a)$$

$$\overleftarrow{h}_i^{d,t} = LSTM(\overleftarrow{h}_{i+1}^{d,t}, v_i, \overleftarrow{\theta}_{LSTM}^{d,t}) \quad (4b)$$

3.5 輸出層

本論文所提出的 E-PGN 模型，採多任務學習，因此在輸出層時同時進行語言模型、命名實體辨識、ABSA、領域分類四種任務結果輸出。

語言模型 LM 語言模型任務預測採用了 NSSoftmax (Negative Sampling Softmax) (Mikolov et al., 2013) 作為預測前後字的模型。前向 LSTM 計算 $x_{1:i}$ 的下一個字 x_{i+1} 的機率，表示為 5a。後向 LSTM 計算 $x_{1:i}$ 的上一個字 x_{i-1} 的機率，表示為 5b。

$$P^f(x_{i+1}|x_{1:i}) = \frac{1}{Z} \exp\{w_{\#x_{i+1}}^\top \vec{h}_i + b_{\#x_{i+1}}\} \quad (5a)$$

$$P^b(x_{i-1}|x_{1:n}) = \frac{1}{Z} \exp\{w_{\#x_{i-1}}^\top \overleftarrow{h}_i + b_{\#x_{i-1}}\} \quad (5b)$$

其中 $\#x$ 代表目標字 x 的在字典中的序號。 $w_{\#x}$ 和 $b_{\#x}$ 分別是目標字的向量和目標字的偏差。 Z 則是通過計算的標準化項。

損失函數則定義為式 6，其中 $D_{LM} = \{(x^n)\}_{n=1}^N$ 。

$$L_{LM} = -\frac{1}{2|D_{LM}|} \sum_{n=1}^N \sum_{t=1}^T \{\log(p^f(x_{t+1}^n|x_{1:t}^n)) + \log(p^b(x_{t-1}^n|x_{t:T}^n))\} \quad (6)$$

命名實體辨識 NER 命名實體辨識的序列預測使用傳統 CRF 模型，將 $\vec{h}_i^{d,t}$ 和 $\overleftarrow{h}_i^{d,t}$ 相加得到 h_i 。因此給定輸入句子 x ，預測標籤序列 $y = l_1, l_2, \dots, l_i$ 的輸出機率 $P(y|x)$ ，如式 7。

$$P(y|x) = \frac{\exp\{\sum_i (w_{CRF}^{l_i} \cdot h_i + b_{CRF}^{(l_{i-1}, l_i)})\}}{\sum_{y'} \exp\{\sum_i (w_{CRF}^{l'_i} \cdot h_i + b_{CRF}^{(l'_{i-1}, l'_i)})\}} \quad (7)$$

其中 y' 表示任意標籤序列， $w_{CRF}^{l_i}$ 是特定於 l_i 的模型參數，而 $b_{CRF}^{(l_{i-1}, l_i)}$ 則為特定於 l_{i-1} 和 l_i 的偏差值。

損失函數定義為式 8， $D_{ner} = \{(x^n, y^n)\}_{n=1}^N$ 。

$$L_{ner} = -\frac{1}{|D_{ner}|} \sum_{n=1}^N \log(p(y^n|x^n)) \quad (8)$$

基於面向情感分析 ABSA ABSA 任務使用 Cross-Entropy 進行情感分類，將 BiLSTM 的

\vec{h} 和 \overleftarrow{h} 相加得到 v_s ，輸入到 softmax，如式 9

$$y = \text{softmax}(W_x v_s + b_s) \quad (9)$$

損失函數如式 10，其中 $\hat{y}_i, y_i \in \{0, 1\}$ 分別表示文本 i 的真實和預測的情感類別。

$$L_{ABSA} = -\frac{1}{N_s} \sum_{i=1}^{N_s} (\hat{y}_i \ln y_i + (1 - \hat{y}_i) \ln(1 - y_i)) \quad (10)$$

域分類器 Domain Classifier 首先將 BiLSTM 輸出的向量視為域分類的文檔，表示為 v_d ，在 v_d 輸入到 Softmax 之前，先通過梯度反轉層。在數學上，將梯度反轉層視為一個偽函數 (pseudo-function)，在反向傳遞時，梯度反轉層從後層獲取梯度 (Gradient)，並且乘上 $-\lambda$ 後，傳遞給前一層。

$$R_\lambda(x) = x \quad (11a)$$

$$\frac{\partial R_\lambda(x)}{\partial x} = -\lambda I \quad (11b)$$

我們可以透過式 11 描述前向和後向傳遞的行為。最後利用式 12 進行預測。

$$d = \text{softmax}(W_d \hat{v}_d + b_d) \quad (12)$$

損失函數如式 13

$$L_{dom} = -\frac{1}{N_s + N_t} \sum_{i=1}^{N_s + N_t} \hat{d}_i \ln d_i + (1 - \hat{d}_i) \ln(1 - d_i) \quad (13)$$

其中 $\hat{d}_i, d_i \in \{0, 1\}$ 分別是文本 i 的真實和預測領域， N_s, N_t 則代表來源域和目標域的資料數量。

整體損失計算 由於 E-PGN 模型採用多任務學習，整體損失函數表示如式 14，將兩個域的三個任務和域分類器的損失，共七個損失函數進行加總。

$$L_{total} = \sum_{d,t} \lambda_t^d L_t^d + \lambda_{dom} L_{dom} \quad (14)$$

因為每個域任務的資料集大小不一，我們使用 DWA (Liu et al., 2019) 來動態地調整不同損失函數間的 λ 大小。在 DWA 中，每個任務先計算前一個 Epoch 對應的損失比，如式 15。

$$w_k(t-1) = \frac{L_k(t-1)}{L_k(t_2)} \quad (15)$$

再將結果代入式 16，得到任務 k 的權重，並除以常數 T ，得到每次 Epoch 任務 k 的權重 λ 。

$$\lambda_k(t) := \frac{k \exp(w_k(t-1)/T)}{\sum_i \exp(w_i(t-1)/T)} \quad (16)$$

t 為 Epoch, T 是一個常數, 代表了任務間的鬆散程度, 如果 T 值越大, λ 會越接近 1, 也就是說各任務間權重差異較小。

4 實驗

本論文使用的資料集共有三種, 分別是使用於目標域的歌手文章, 來源域的新聞、家電評論, 以下將分別介紹這三個資料集的訓練與測試資料數量及標記策略。

歌手文章 歌手文章是本論文的目標域資料集, 擷取自『批踢踢 PTT 實業坊』論壇與音樂相關的 97 個版, 標記方式參考 Chiu (2020) 所使用的資料集, 標記人員會依照段落的敘述, 來判斷該段落的實體, 以及對實體表達的情感極性, 情感極性有三種類別, 分別為正向、負向、中立, 資料比例分別為 28%、22%、50%, 詳如表 1。

	段落數量	實體		情感		
		歌手	正向	負向	中立	
訓練資料	4,000	6,715	1,834	1,475	3,406	
測試資料	1,000	1,710	489	382	839	

Table 1: 歌手文章的訓練與測試資料集

新聞 新聞資料作為來源域的資料集, 針對各種不同的新聞網站, 將提及人名實體的文章進行蒐集。標記方式參考 Chou et al. (2016) 所使用的資料集, 是透過自動標記的方式, 一共標記了訓練和測試資料分別 15,000 句與 500 句, 共 20,000 句, 如表 2。

	句子數量	人名實體個數
訓練資料	15,000	3,723
測試資料	500	155

Table 2: 新聞的訓練與測試資料集

家用電器評論 家用電器評論的資料集作為本論文的另外一個來源域, 取自『Mobile01』與家用電器有關的五個版, 使用者的發文及文章底下的討論串, 作為我們的資料集。標記方式參考 Kan (2021) 的人工標記, 共使用了 2,000 個句子的家電評論, 其中訓練和測試資料分別為 1,500 句與 500 句, 具有產品廠牌 (Product Name, PN) 以及產品名稱 (Brand Name, BN) 的情感標記, 情感類別包括正向、負向、中立等三種, 如表 3。

效能評估與參數設定 在命名實體辨識任務上, 我們採用完全比對 (Exactly Match) 的方式來進行評估。ABSA 任務亦同, 也就是情感

	句子數量	實體		情感		
		PN	BN	正向	負向	中立
訓練資料	1,500	2,075	2,538	968	484	186
測試資料	500	650	820	373	187	260

Table 3: 家電評論的訓練與測試資料集

標記的數量、極性必須和實體完全一致, 才會納入計算。

表 4 為本論文 E-PGN 模型的參數設置, 在這個設定下, 可以得到最佳的效能。

#of words per sentence	128	Batch size	30
Word Embedding	100	Dropout	0.5
Task embedding	8	Learning rate	1e-3
BiLSTMLayer	2	L2-regularization	1e-8
Hidden	200	Optimizer	SGD

Table 4: 參數設定

標記策略 在命名實體辨識任務上, 我們參考 AutoNER (Shang et al., 2018) 的 Tie/Break 的標記方式。並使用 CKIP Tagger (Li et al., 2020) 對中文資料進行斷詞。

Token	吳	亦	凡	倒	數	與	你	的	時	間	距	離
Chunk	B	T	T	B	T	B	B	B	B	T	B	T
Type		A		N	N	N	N	N	N		N	
Sentiment	Neu	Neu	Neu	N	N	N	N	N	N	N	N	N

Figure 2: 標記範例

圖 2 中, Chunk 為 CKIP 斷詞後的結果, B 表示 Break, 與前字無關, T 表示 Tie, 可與前字合為一個詞。Type 代表是否為實體, A 表示 Artist, N 表示 None。Sentiment 代表實體的情感極性, 分為正向 (Pos)、負向 (Neg)、中立 (Neu)、無 (N)。

模型效能 本論文的目標域為歌手文章, 依據來源域的不同, 分成兩個部分 (家用電器評論、新聞), 並在不同的來源域的實驗下, 再分別進行輸入資料為兩個任務 (LM, NER) 以及三個任務 (LM, NER, ABSA) 的實驗。為了提高實驗的準確性, 每個實驗皆執行五次後計算平均值。

4.1 來源域為家用電器評論

兩個任務 (LM, NER) 這個實驗使用的輸入資料只有命名實體辨識和語言模型兩個資料集, 並進行 BIESO 與 Tie/Break 標記的比較。

圖 3 可以看到, 使用 Tie/Break 標記可以得到 0.85 的 F1, 相較於 BIESO 標記效能 (0.83) 相差了 2%, 但和 BaseLine (Chiu, 2020) 相比, 還有一點成長的空間。從結果可以得知

使用 Tie/Break 標記，有助於模型效能，因此後續實驗皆使用 Tie/Break 的標記方式。

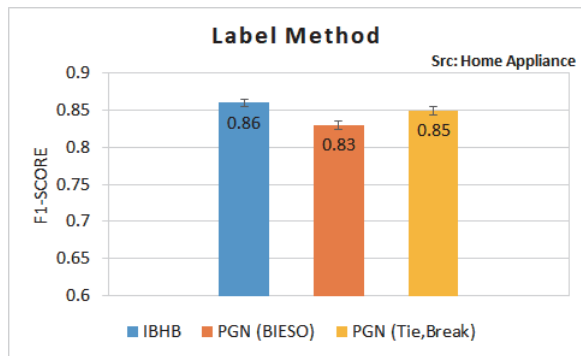


Figure 3: BIESO 和 Tie/Break 比較 (src: 家電評論)

三個任務 (LM, NER, ABSA) 這個實驗新增了 ABSA 任務，因此輸入資料有三個任務的資料集，並比較固定 (Fixed) 和動態 (DWA) 的兩種計算 Loss 權重的方式。

圖 4 顯示，在命名實體辨識任務上，較沒有兩個任務時的效能來的好，和 Baseline 相比，Fixed 和 DWA 皆略有下降。但在 ABSA 任務上，不論是 Fixed 或是 DWA，均顯著優於 BaseLine，且使用 DWA 優於 Fixed，在三種情緒的平均值分別相差了 8%、18%。另外也可以從圖中看出來，使用 DWA 的五次實驗的標準差也比 Fixed 的標準差小。另外不論是 Fixed 或是使用 DWA，兩個不同的實驗，五次實驗的結果差距皆不大，只相差了 0.06。因此在後續的實驗中，皆以 DWA 的計算方式來進行。

4.2 來源域為新聞

兩個任務 (LM, NER) 在來源域有大量的標記資料下，可以從圖 5 看到，BIESO 和 Tie/Break 標記的效能皆優於 BaseLine，相差了 2% 和 4%。因此根據實驗結果，在 PGN 來源域的資料多於目標域資料時，可以得到不錯的效能。

三個任務 (LM, NER, ABSA) 加上 ABSA 任務後，比較使用領域嵌入和梯度反轉層的效能，圖 6 所示。在命名實體辨識任務上僅有微小的差距，E-PGN 的 F1 達到 85.5%，PGN 則是 85.2%，雖然在命名實體辨識任務上沒辦法優於 BaseLine 的效能 87.4%，但在 ABSA 任務有顯著的成長，PGN 的 SA 平均 F1 為 70%，E-PGN 為 78%，和 BaseLine 的效能 56% 相比，有大幅的成長。因此在領域的處理上，和領域嵌入相比，使用梯度反轉層對模型有明顯的幫助。

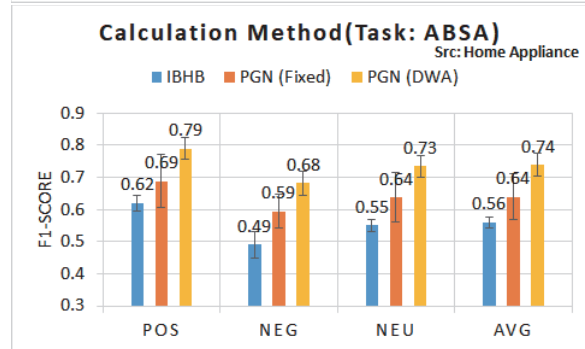
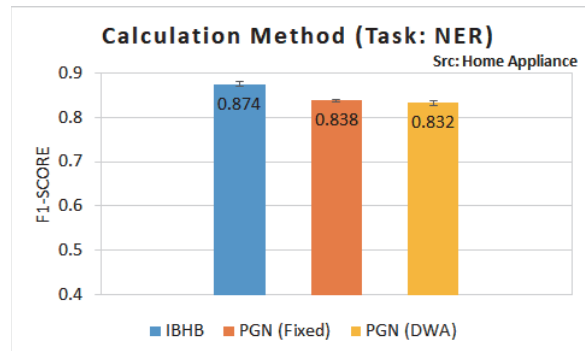


Figure 4: Fixed 和 DWA 比較 (src: 家電評論)
 (上)NER 任務 (下)ABSA 任務

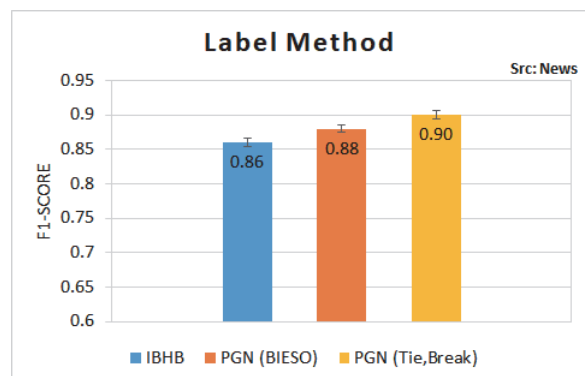


Figure 5: BIESO 和 Tie/Break 比較 (src: 新聞)

詞嵌入 (Word Embedding) 比較 本實驗比較了使用 GloVe 和 BERT 的結果，圖 7 可以看到，不論在 ABSA 或是命名實體辨識任務上，使用 BERT 時，效能皆低於使用 GloVe。因此認為模型在使用 BERT 時，會產生太多的參數，反而造成模型無法得到較好的效能。

5 結論

為了有效辨識社群網路上歌手的網路聲量，本論文參考了參數生成網路 (Jia et al., 2019) 與 DANN 模型 (Ganin and Lempitsky, 2015)，提出擴展參數生成網路 (E-PGN) 的遷移學習方法來達成跨領域的命名實體辨識和 ABSA 任務，並解決目標域訓練資料不足的問題。由

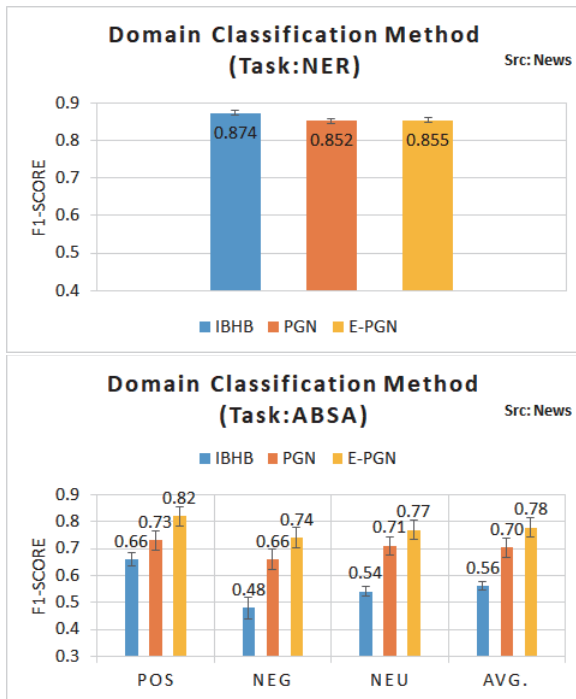


Figure 6: Domain Emb. 和 GRL 比較 (src: 新聞)
 (上)NER 任務 (下)ABSA 任務

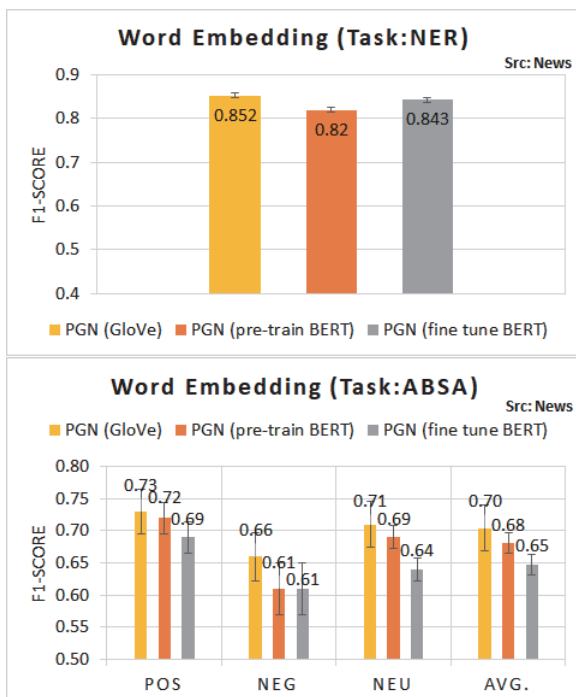


Figure 7: Word Emb 比較 (src: 新聞)
 (上)NER 任務 (下)ABSA 任務

於採用多任務學習 (Multi-task Learning) 進行命名實體辨識和 ABSA 的任務，因此在訓練過程中會產生大量的參數。不同於原 PGN 模型，我們移除了領域嵌入，並改用了梯度反轉層 GRL 進行領域分類，並且在計算整體損失函數的部分引用動態調節權重的機制 (Liu

et al., 2019)，有效進行權重設定。

實驗結果顯示，使用 Tie/Break 標記可以提升模型辨識效能；在權重設定的比較中，使用 DWA 動態設定權重，確實能有效提升模型效能。在家電評論作為來源域的實驗中，雖然家電評論只有少量資料，但模型效能仍可在 ABSA 任務上有顯著的成長。在目標域資料不足下，本論文展現了 E-PGN 模型較原 PGN 仍能獲得良好的效能。

References

- Wei-Cheng Chiu. 2020. Joint learning of aspect-level sentiment analysis and singer named recognition from social networks. In *International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*.
- Chien-Lung Chou, Chia-Hui Chang, and Ya-Yun Huang. 2016. Boosted web named entity recognition via tri-training. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 16(2).
- Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, page 1180–1189. JMLR.org.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 504–515.
- Mengting Hu, Yike Wu, Shiwan Zhao, Honglei Guo, Renhong Cheng, and Zhong Su. 2019. Domain-invariant feature distillation for cross-domain sentiment classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5558–5567.
- Chen Jia, Xiaobo Liang, and Yue Zhang. 2019. Cross-domain NER using cross-domain language modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2464–2474. Association for Computational Linguistics.
- Tai-Jung Kan. 2021. Home appliance review research via adversarial reptile. Master's thesis, National Central University.
- Peng-Hsuan Li, Tsu-Jui Fu, and Wei-Yun Ma. 2020. Why attention? analyze bilstm deficiency and its remedies in the case of ner. In *AAAI*.

- Zheng Li, Xin Li, Ying Wei, Lidong Bing, Y. Zhang, and Qiang Yang. 2019. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning.
- Shikun Liu, Edward Johns, and A. Davison. 2019. End-to-end multi-task learning with attention. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1871–1880.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’13, page 3111–3119. Curran Associates Inc.
- Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543.
- Jingbo Shang, Liyuan Liu, Xiang Ren, X. Gu, Teng Ren, and Jiawei Han. 2018. Learning named entity tagger using domain-specific dictionary. In *EMNLP*.
- Kai Zhang, Hefu Zhang, Qi Liu, Hongke Zhao, Hengshu Zhu, and Enhong Chen. 2019. Interactive attention transfer network for cross-domain sentiment classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):5773–5780.