The 2016 Conference on Computational Linguistics and Speech Processing
ROCLING 2016, pp. 115-128

運用序列到序列生成架構於重寫式自動摘要

Exploiting Sequence-to-Sequence Generation Framework for Automatic Abstractive Summarization

謝育倫 Yu-Lun Hsieh, 劉士弘 Shih-Hung Liu, 陳冠宇 Kuan-Yu Chen,
王新民 Hsin-Min Wang, 許聞廉 Wen-Lian Hsu
中央研究院資訊科學研究所
{morphe, journey, kychen, whm, hsu}@iis.sinica.edu.tw

陳柏琳 Berlin Chen
國立臺灣師範大學資訊工程學系
berlin@ntnu.edu.tw

摘要

自動摘要 (Automatic Summarization) 一直以來都是熱門的研究議題，過去多著重在節錄式 (Extractive) 摘要，而重寫式 (Abstractive) 摘要相當稀少。有鑑於近期深度學習被廣泛應用在自然語言處理，尤其是機器翻譯等領域的成功，讓重寫式摘要的研究又熱絡起來。近期文獻中已初步驗證了遞歸神經網路 (Recurrent Neural Network) 在文件的重寫式自動摘要之成效。因此本文欲探討加入注意力 (Attention) 機制的效果。注意力機制的特點是它能夠在生成文字的同時，對於關鍵片段增強注意力，藉此產生更佳的摘要。此外本文亦欲探究單向 (Uni-directional) 及雙向 (Bi-directional) 遞歸神經網路的差異。本文採用語料是大規模中文短文摘要集 (Large-scale Chinese Short Text Summarization Dataset, LCSTS)。結果顯示，本文所提出之改進對於摘要品質有明顯的助益。

關鍵詞：重寫式自動摘要、序列到序列、遞歸神經網路

一、緒論

隨著大數據時代的來臨，巨量的文字訊息充斥於網際網路之中，並且被快速地傳遞並分享於全球各地，資訊超載 (Information Overload) 的問題也因此產生。如何能讓人們快速且有效率地瀏覽或消化與日俱增的資訊，已成為一個刻不容緩的研究課題，其中自動摘要 (Automatic Summarization) 更是不可或缺的關鍵技術 [1]。自動摘要之目的在於擷取單一文件 (Single-Document) 或多重文件 (Multi-Document) 中的重要語意與主題資訊，讓使用者能更有效率地瀏覽與理解文件的主旨，並快速地獲得其中關鍵資訊，省去

大量審視文件時間。

約略來說，自動摘要研究可分為二大類，節錄式(Extractive)摘要與重寫式(Abstractive) 摘要（或稱抽象式摘要）。前者主要是依據特定的摘要比例，從最原始的文件中選取重要的語句來組成摘要；而後者是在完全理解文件內容之後，重新撰寫產生摘要來代表原始文件的內容，其所使用之詞彙不全然來自於原始文件。此種摘要方式可說是最貼近人們日常撰寫摘要的形式。然而，重寫式摘要需要複雜的自然語言處理(Natural Language Processing, NLP) 技術，如資訊擷取 (Information Extraction)、對話理解 (Discourse Understanding) 及自然語言生成 (Natural Language Generation) 等 [3][4]，因此，過去主流的研究仍著重在節錄式自動摘要 [5]。

近年來我們可以看到深度學習 (Deep Learning) 方法被廣泛應用在各大領域，並取得相當不錯的效果[6][7][8][9]。其中，序列到序列 (Sequence-to-Sequence) 生成架構更是在機器翻譯領域中獲得相當耀眼的成果[10][11][31]，並已初步被應用到重寫式自動摘要之研究上[12][13][14][15]，本論文將延續此一主軸，進而提出兩個研究貢獻。第一，基於序列到序列生成架構，本論文提出利用注意力 (Attention) [30] 機制來增進生成架構之模型，由於此模型會對於關鍵片段增強注意力，使得在生成文字摘要時更能含括原始文件中的重點主題或語意。第二，本論文亦欲探索遞歸神經網路模型的雙向(Bi-direction) 建模方法，利用此法可以更完整的捕捉序列中各單位間的相關性，使得所生成的序列能夠更有效地代表原文的內容，藉以增進重寫式自動摘要之效能。

本論文後續安排如下：第二節扼要地介紹現今自動摘要技術的相關研究與發展；第三節首先介紹基本的遞歸神經網路原理，然後闡述如何使用序列到序列生成架構來自動產生出文字摘要，並且說明如何藉助注意力機制來改進序列到序列的生成模型，使其產生的文字內容得以更精準地代表原始文件的內容；第四節介紹實驗語料與設定以及摘要評估之方法；第五節說明實驗結果及其分析；第六節為結論與未來研究方向。

二、自動摘要技術

1、節錄式摘要技術

我們將過去節錄式摘要研究所陸續發展出的技術大略地歸納成兩大類 [2]：

a. 以非監督式機器學習為基礎之自動摘要模型技術：非監督式機器學習通常將自動摘要任務視為一排序並挑選具代表性語句之問題。其核心方法通常是計算出一種或數

種特徵值以供語句排序使用，其中常見的特徵有：語句與文件相關性 [16]、語句所形成的語言模型生成文件之機率 [17]、語句間相關性 [18][19]、或語句與文件在潛藏主題空間中的距離關係 [20] 等。

b. 以監督式機器學習為基礎之自動摘要模型技術：監督式機器學習通常將自動摘要之任務視為一個二元分類 (Binary Classification)，亦即將語句區分為摘要或非摘要語句。在訓練這樣的分類器前，必須事先準備好一些訓練文件以及其對應的人工標註摘要資訊，然後透過各種分類器的學習機制進行模型訓練。接著對於尚未被摘要之測試文件，將裡頭的每個語句進行二元分類，即可依其結果產生出摘要。此類方法中較著名的包括簡單貝氏分類器 (Naïve Bayes Classifier) [21]、高斯混合模型 (Gaussian Mixture Model, GMM) [22]、隱藏式馬可夫模型 (Hidden Markov Model, HMM) [23]、支援向量機 (Support Vector Machines, SVM)、及條件隨機場域 (Conditional Random Fields, CRF) [24]等。監督式學習可同時結合多種摘要特徵值來表示每一語句（包含上述以詞彙或結構為基礎之摘要方法，以及各式非監督式模型針對語句所輸出的分數或機率值），以做為監督式摘要模型判斷語句是否屬於摘要語句的依據 [20]。

2、重寫式摘要技術

近年來重寫式自動摘要技術均基於深度類神經網路 (Deep Neural Network) 之方法來達成，尤其是序列到序列生成架構在機器翻譯領域取得成功後[11]，許多學者採用類似的架構來應用到重寫式摘要這個研究中，如 [12] 直接套用機器翻譯的序列生成模型來作重寫式自動摘要，頗有成效，驗證了此一方向的可行性。另一方面，原始文件中的重要資訊若含有專有名詞（例如：人名、組織名、地名等），則產生的摘要也應合理地將其放入摘要中，基於此想法，[14] 提出了一個複製機制 (Copy mechanism) 使重要的專有名詞能夠正確地包含在自動摘要文字中，以產生內容更加豐富的摘要。最後，也有學者提出分散注意力 (Distraction) 機制 [15] 來應用到序列到序列生成模型上，其主要概念是希望產生出來的摘要內容能涵蓋更多原始文件的主題或面向，並且期望能藉此來避免冗餘的資訊。本論文的研究主要是基於 [12] 的架構做改良，除了提出使用直觀的注意力機制並加入雙向建模方式來改善，另外本論文所使用的遞歸神經網路中的類神經元 (Cell) 也不同於原始論文中的閘循環單元 (Gated Recurrent Unit, GRU) [26]，我們採用的是長短期記憶 (Long Short-Term Memory, LSTM) [27] 來建構遞歸神經網路，達到重

寫式自動摘要之目的。

三、序列到序列生成架構

在本節中，首先介紹遞歸神經網路 (Recurrent Neural Network, RNN) 及其演進，還有如何加入雙向建模方法來學習更強健的模型；接著說明序列到序列的生成架構如何運用於重寫式摘要中，最後介紹注意力機制來增進序列到序列生成架構。

1、遞歸神經網路

現今泛用的遞歸神經網路雛形早在 1980 年代就有人提出[28]，和前饋神經網絡 (Feed-forward Neural Network, FNN) 最主要的差異在於，遞歸神經網路可以用來學習一個序列的資訊。關鍵在於，其包含一隱藏的狀態層 (State layer)，用來儲存歷史資訊，可類比於人腦中的記憶。一個最基本的遞歸神經網路運作如下：當接受到序列的一個輸入值時，此狀態層的內容會根據歷史以及現有的輸入，來決定下一個時間點的狀態為何。以數學式定義來說，令輸入層為 x，輸出為 y，狀態層 s，則在序列的時間點為 t 時，輸入表示為 $x(t)$，狀態為 $s(t)$，而輸出則為 $y(t)$。那麼在此時間點網路內各層的計算可以下式表達：

$$s(t) = \sigma\big(W \cdot x(t) + U \cdot s(t-1)\big), \tag{1}$$

$$y(t) = g(V \cdot s(t)), \tag{2}$$

其中，$\sigma(*)$ 代表的是 S 形 (Sigmoid) 函數，而 $g(*)$ 代表的是軟性最大 (Softmax) 函數，W, U, V 為權重矩陣。當此模型應用在文字處理的時候，通常輸入的值為字向量 (Word embeddings) [29]。而輸出通常為一個維度等於字彙個數的向量，其代表的意義為某個字出現的機率分布。

然而，此基本遞歸神經網路存在一些限制，最明顯的即為梯度消失問題 (Gradient Vanishing)。因此，Hochreiter 與 Schmidhuber [27] 提出了長短期記憶 (LSTM) 這個單元來建立遞歸神經網路，以避免上述問題。LSTM 比簡單遞歸神經網路中的單元複雜許多，但其核心概念為，利用閘 (Gate) 這個機制來限制隱藏的記憶層及輸入輸出資訊

量。更深入來說，LSTM 架構中含有三個閘：輸入閘 (Input gate) 、輸出閘 (Output gate)、及遺忘閘 (Forget gate)，並有一記憶單元 (Memory cell)。它們恰如其名的分別代表了三種不同的資訊流控制，以及所謂的「記憶」機制。其詳細數學定義為：三個閘則為 i, o, f，分別代表輸入，輸出，及遺忘。在序列的時間點為 t 時，輸入表示為 $x(t)$，隱藏層為 $h(t)$，記憶為 $C(t)$。各閘分別定義如下：

$$f(t) = \sigma\big(W_f \cdot [h(t-1), x(t)]\big), \tag{3}$$

$$i(t) = \sigma(W_i \cdot [h(t-1), x(t)]), \tag{4}$$

$$o(t) = \sigma(W_o \cdot [h(t-1), x(t)]), \tag{5}$$

其中 W_* 代表各閘所對應的權重矩陣。而記憶層是由輸入及遺忘閘來控制，定義如下：

$$\tilde{C}(t) = tanh(W_C \cdot [h(t-1), x(t)]), \tag{6}$$

$$C(t) = f(t) \cdot C(t-1) + i(t) \cdot \tilde{C}(t), \tag{7}$$

其中 $\tilde{C}(t)$ 代表的是記憶層的候選值。最後所得的輸出則為：

$$h(t) = o(t) \cdot tanh(C(t)) \tag{8}$$

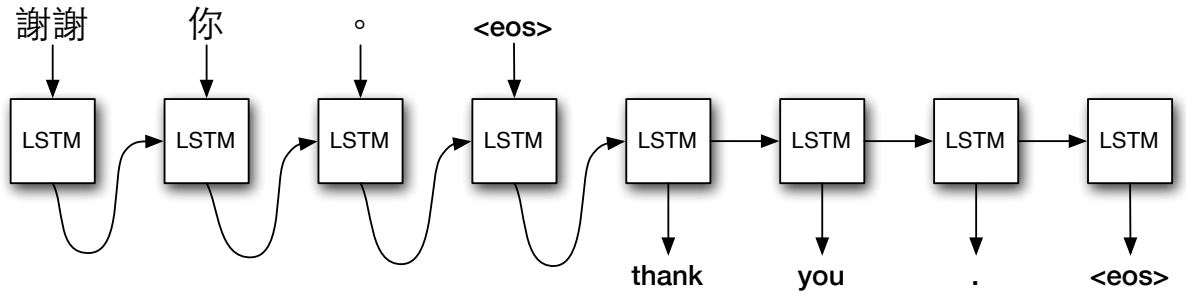
由以上定義可知，LSTM 藉由輸入及遺忘閘來控制記憶層中所儲存的資訊，配合輸出閘來調節輸出的權重，可以避免梯度消失問題，達到學習較長序列的效果。

至於建立雙向遞歸神經網路模型，其實是一個非常直觀的改進。我們僅需將輸入的序列反轉過來，並套用同樣的架構，即可得到另一個反向的序列資訊向量。最終，正向和反向的兩個向量，可以以串接 (Concatenate) 或者加總等方式合併起來，即成為最終的輸出值。這樣做的目的在於學習到一個字詞的左側及右側語意資訊，藉以達到產生更佳摘要的效果。

2、序列到序列模型

序列到序列模型是遞歸神經網路的其中一種延伸，又稱為「編碼—解碼器」(Encoder-Decoder)。其核心概念為利用遞歸神經網路來學習一個序列的所有資訊（或稱

為「編碼」)，並將之濃縮至一個向量中，再利用另一個遞歸神經網路來將此資訊「解碼」出來，進而生成另一個序列，故得此名。以常見的機器翻譯為例，假設我們希望將一個中文語句翻譯成英文，那麼可以利用遞歸神經網路來先將中文句子中每個字的資訊都學習到一個向量中，再利用它生成另一個序列的英文句子，即為我們翻譯的目標。經過巨量中英對照的語料訓練之後，一個序列到序列模型將可以自動學習到中英文句間的對應關係，達到翻譯的效果。序列到序列模型可以用圖一中的範例來說明，首先將一中文句子中的每個詞依序當成輸入送到遞歸神經網路中，另外最後輸入一特殊符號 “<eos>”，當此網路接收到此特殊符號時，即代表編碼完成，可以進入解碼階段。此時，LSTM 的記憶層中，所儲存的即為整個句子中的所有重要資訊，可以用來依序解碼出正確的英文翻譯句。另外，在生成的時候常配合使用束搜尋 (Beam search) 來改善結果。特別注意的是，在圖一中的多個 LSTM 單元實際上是同一個，如上節所述，遞歸神經網路的特點是可以同時學習歷史資訊和當前資訊，故圖一呈現的可以看成是一個展開後 (Unrolled) 的遞歸神經網路。



圖一、序列到序列模型應用於機器翻譯示意圖

3、注意力機制

注意力機制是用來改進在上述模型中生成階段的效果。其實質作法為，在生成階段時，額外學習一組注意力權重，代表著目前生成的字和輸入序列間各字的相關性。再以機器翻譯為例，一種常見的狀況是，來源和目的語言間詞彙和文法的對應為非線性，比如其中一個語言可能將動詞放在句首，另一個則放在句尾。這個時候，注意力機制可以學習到序列間各單元的對應關係，藉以達到更精確的翻譯效果。若應用在重寫式摘要這個工

作，我們可以預期在生成摘要的某個部分時，注意力機制將會幫助模型選擇高度相關的原始文句語意，進而產生出更好的摘要。注意力機制的詳細運作如下。首先定義在編碼階段時間 t 的注意力向量 (Attention vector) $a(t)$ ，如下所示 [31]：

$$a(t) = \frac{\exp\left(h_S(t)^{\mathrm{T}} \cdot W_a \cdot h_D(t)\right)}{\sum_{t'} \exp\left(h_S(t')^{\mathrm{T}} \cdot W_a \cdot h_D(t')\right)}, \tag{9}$$

其中 $h_S(t)$ 代表在解碼階段時間 t 的 LSTM 輸出值，而 $h_D(t)$ 代表編碼階段時間 t 的 LSTM 輸出值，W_a 為注意力權重矩陣。有了注意力向量後，我們定義內容向量 $c(t)$ 為

$$c(t) = h_D(t) \cdot a(t), \tag{10}$$

其所代表的意義為在時間 t 時經過注意力向量加權後的內容特徵值。此內容向量 $c(t)$ 與時間 t 的 LSTM 隱藏層 $h_S(t)$ 合併後，再通過 Softmax 函數，即可依序生成摘要文字。這樣一來，我們就可以利用在不同時間點，經不同注意力篩選過後的內容，以利在生成摘要時，更容易擷取到原始內容的重要資訊。

四、實驗語料及評估方法

1、實驗語料

本論文實驗語料庫為公開的大規模中文短文摘要集(Large-scale Chinese Short Text Summarization Dataset, LCSTS) [12]，是由新浪微博網站所收集而來，內容均為新聞報導，並額外經過多人以 1 ~ 5 分來評估其摘要品質，之後挑選經 3 人標注後均達 3 分以上的摘要作為測試集，以確保測試資料的可靠度。本文採用和 [12] 相同的標準訓練及測試集切分方法，以便與相關研究作合理的比較[1]。詳細的語料庫統計資訊如表一所示。舉例來說，其中一篇文章和其摘要如下（已由簡體轉為繁體）：

文：	水利部水資源司司長陳明忠今日在新聞發佈會上透露，根據剛剛完成的水資源管理制度的考核，有部分省接近了紅線的指標，有部分省超過紅線的指標。在一些超過紅線的地方，將對一些取用水項目進行區域的限批，嚴格地進行水資源論證和取水許可的批准。
摘：	部分省超過年度用水紅線指標 取水項目將被限批

[1] 訓練集其中 591 篇為驗證集 (validation set)

表一、實驗語料統計資訊

	訓練集	測試集
文件數	2,400,591	725
文件平均字元數	103.7	108.1
摘要平均字元數	17.9	18.3

2、系統設定

在本文的實驗中，語料是用字元的形式送入模型來學習，也就是不經過斷詞，因為據 [12] 結果顯示字元效果較佳。字元個數上限為語料庫中前 4,000 個最常出現的字，與前人研究一致。程式部分是基於 Torch 深度學習工具建置[2]。我們使用一層遞歸神經網路搭配注意力機制，並比較單向及雙向建模法的差異，以及其中的 LSTM 單元維度、字向量維度等參數所帶來的影響。其餘不變的參數設定為：最佳化 (Optimization) 方法使用梯度下降法(Stochastic Gradient Descent, SGD)，學習率 (Learning rate) 為 1，訓練回數 (Epoch) 最多為 20 回，在超過十回後，學習率每回將以 90% 遞減。梯度範數 (Gradient norm) 上限設為 5。在有使用圖形處理單元 (Graphic Processing Unit, GPU) 加速的環境下，一種參數組合完整訓練時間約為 48 小時。

3、成果評估

本文採用的評估方法為自動摘要最常用的「召回率導向的摘要評估」(Recall-Oriented Understudy for Gisting Evaluation, ROUGE) [25]。ROUGE 方法是計算自動摘要結果與答案之間的單位重疊量占參考答案總單位數的比例，這邊所使用的單位可以是 *N*-連詞 (*N*-gram) 或者詞序列 (Word Sequence)，一般常使用的是最長相同詞序列 (Longest Common Subsequence)。本文使用三種評估方式：ROUGE-1 (Unigram)、ROUGE-2 (Bigram) 和 ROUGE-L (Longest Common Subsequence) 分數。直觀來看，ROUGE-1 可以說是代表自動摘要的訊息量，ROUGE-2 則是評估自動摘要的流暢性，而 ROUGE-L 可看成是摘要對原文的涵蓋率。因本研究著重在產生重寫式摘要，故摘要的流暢性是一個相當重要的指標，因此實驗數據以 ROUGE-2 分數為觀察重點。為增進易讀性，以下數據呈現簡寫為 R-1、R-2 及 R-L。

[2] http://torch.ch。實際運作部分程式改自 https://github.com/harvardnlp/seq2seq-attn

表二、維度及方向性對摘要品質影響實驗結果

方向	字向量維度	RNN 維度	R-1	R-2	R-L
單	128	128	0.305	0.188	0.280
單	128	300	0.328	0.206	0.303
單	300	128	0.315	0.193	0.285
單	300	300	0.348	0.222	0.320
雙	128	128	0.324	0.207	0.305
雙	128	300	0.360	0.235	0.335
雙	300	128	0.332	0.213	0.311
雙	300	300	**0.369**	**0.243**	**0.343**

另外，本文也比較了其他使用同一語料的研究，以便觀察本文所提出的模型與現有方法成效之差異。首先是一個使用 GRU 的單向遞歸神經網路序列到序列模型 [12]。另外一個則是據作者所知，以此語料為實驗對象的文獻中，最先進 (State-of-the-art) 的方法 [15]，其中同樣使用 GRU 建立雙向、多層遞歸神經網路並內含分散注意力機制的序列到序列模型。

四、實驗結果與討論

1、維度及方向性實驗

本實驗旨在測試維度及方向性等參數對於摘要品質的影響，本文測試了字向量維度為 128 或 300，以及遞歸神經網路中 LSTM 各單元維度為 128 或 300，還有使用單向或雙向網路等不同設定，對於重寫式摘要品質所造成的影響。另外需補充說明，因硬體限制，我們無法再進一步提升單元維度。

結果如表二所示，無論使用單向或雙向，維度加大可以帶來相當的助益。這代表在學習序列資訊這個工作中，神經單元維度越高，越能夠呈載更豐富的資訊。但我們也可以發現，若遞歸神經網路中 LSTM 單元維度較低時，提升字向量的維度並不會帶來太大的幫助。另外，在同樣的維度設定下，使用雙向網路的效果均較單向佳，這顯示了在產生重寫式摘要時，需考慮到原始文章中，前後文統整後的內容，而不能僅靠單方面的資訊。總歸來說，使用足夠維度的神經單元配合雙向網路，可以達到最佳效果。

表三、本文方法與前人研究比較結果（R-1、R-2 及 R-L 均為 F-score 值）

方法	方向	字向量維度	RNN 維度	R-1	R-2	R-L
HU	單	-	-	0.299	0.174	0.272
CHEN	雙	500	500	0.352	0.226	0.325
本文	雙	300	300	**0.369**	**0.243**	**0.343**

2、與前人研究比較

本節比較上述參數設定實驗中，表現最佳的結果，以及兩種前人提出的方法。首先是 [12] 所提出來的方法，其中所採用的單向網路是不含注意力機制的（簡稱為 HU），以及另一個雙向多層網路並含分散注意力機制的 [15]（簡稱為 CHEN）。特別注意的是，CHEN 的方法與本研究所提出的架構相比，除了神經網路較深及維度較高之外，其分散注意力機制也較為複雜，並且所採用的單元為 GRU，不同於本文所使用的 LSTM。其餘技術細節請見原出處的說明，在此不多贅述。

由表三結果，我們首先可以知道，CHEN 的雙向遞歸神經網路模型，相較於 HU 的基線 (Baseline) 方法，已經有大幅度的改善。這再次顯示雙向網路的確可以有效的學習到更多資訊。至於本研究所提出的模型，使用比 CHEN 的方法更為簡單的注意力機制，搭配較低維度的遞歸神經網路，即可在三個不同的評估指標上，都超越現有的最佳結果成效，特別是在 R-2 這個指標上，有超過 2% 的進步。這個現象可能代表，當維度過高時，產生了過度擬合 (over-fitting) 的現象，導致所產生的摘要與答案差異變高。另外，也可能是因為所採用的單元結構不同所致。總而言之，經由本研究所提出的方法所產生出來的摘要，不管是內容涵蓋率或是可讀性，均有明顯的提升，可說是更加接近人類所撰寫的文字。

3、重寫式自動摘要之質性分析

本節旨在以實際舉例的方式來展示本文所提出的模型所產生的自動摘要，以利讀者作主觀的質性分析，並檢驗內容的正確性及可讀性等，難以使用量化分析指標 ROUGE 評估的特性。首先我們將所產生的摘要依照 R-2 的分數由高到低排序，之後選出數則最高（R-2 分數高於 0.8）及最低（R-2 分數為 0）的例子以供討論。

首先，表四中列舉 R-2 分數高於 0.8 的摘要及原文。可以看出，本文所提出的方法

表四、品質優良的自動重寫式摘要與原文及答案對照比較結果

文：	正處於風口浪尖的國內奶粉行業出現大交易。蒙牛乳業（02319.HK）以及雅士利（01230.HK）昨日發佈公告稱，蒙牛乳業將斥資 81.4 億港元收購雅士利約 65.4%股權。業界稱，此舉有助於蒙牛乳業補上奶粉短板，以期重新超越伊利成為行業領頭羊。
摘：	蒙牛 81.4 億港元收購雅士利
答：	蒙牛 81 億港元收購雅士利
文：	27 日，六名全國人大代表聯名向全國人大發出建議書，建議取消徵收社會撫養費。建議書認為，徵收社會撫養費，是把「提倡」變成了「強制要求」，侵犯了公民的合法權益。根據不完全統計，全國每年徵收的社會撫養費超過 200 億元。
摘：	全國人大代表建議取消社會撫養費
答：	六位全國人大代表聯名建議取消社會撫養費

表五、品質差的自動重寫式摘要與原文及答案對照比較結果

文：	症狀表現：頭疼胸悶、手心出汗，心緒不寧。常會自言自語：到底選哪款好，怎麼辦，幾款我都好喜歡！小編認為，最重要的第一步是狠狠地深呼吸，除了為使在大戰中頭腦清醒外，還能順便提前「倒吸一口涼氣」，因為您的錢包又要被掏空了。
摘：	你的錢包好喜歡
答：	雙十一攻略：當網購狂遇上「選擇困難症」時
文：	李克強此次東歐之行，為中國與中東歐國家傳統友誼的延伸鋪路架橋，為雙方互利共贏的經貿合作穿針引線，為中歐戰略夥伴關係的全面發展添火加柴。經過三年「16+1」機制的運作，當前中國與中東歐的合作成效初顯。
摘：	李克強與中東歐合作初顯
答：	地理上的「遠親」心靈上的「近鄰」

可以生成易懂且涵蓋重點內容的摘要，但可能是因為有細節被原始摘要所省略（如第一例中的 “81.4” 和 “81”）或者被自動學習的模型給省略（如第二例中的 “六位”、“聯名”），因此導致 R-2 分數略低，但可說是無損於理解原文中的重點資訊。這些例子顯示，本文提出的重寫式摘要模型有相當的成效，的確可以產生出如同真人所撰寫的摘要內容。

另一方面，表五中列出一些被評估為品質低的例子（R-2 分數為 0）。首先我們可以看到此模型有可能會產生出不合理的摘要文句（如表五第一個例子），雖仍屬於在原始文章中有出現過的字眼，但無法理解其語意。此種錯誤在機器生成語言的時候相當常見，我們推測是因為在進行束搜尋時所依據的是總體機率，並未考慮到文法等語言特性所致。未來可考慮將句法及語意合理性等特徵融合到模型訓練中，以避免此類問題。此外，我們也可以看到本例中摘要長度過短，這是因為我們設定在生成時只要發現所生成的字結

尾為“eos”（也就是句尾符號）即停止，目的是為了避免產生不完整的句子，但可能也因此導致了過短的現象。未來可以改進生成機制，不以句尾符號為停止條件，而改由加入長度限制，以盡可能的生成較長的摘要，達到更好的效果。接下來，我們由表五的例二可以觀察到另一種情形為：自動摘要模型生成結果雖屬合理，但因人們在寫作時偶而會有更深層的譬喻、引申等技巧，而使用了與原始文章完全不相同的文字。以目前的技術來說，這仍屬於一個相當困難的問題，也顯示了在自動重寫摘要方面，還有許多研究發展的空間。

六、結論與未來方向

本論文提出基於遞歸神經網路的深度學習方法，在重寫式摘要這個工作上有相當的成果，且簡化了前人所提出相對較複雜的模型。相比於節錄式摘要，此方法能產生更完整，並且讓人更能快速掌握文章中的重點資訊的摘要。未來，我們計畫有許多改進方向，例如引入事前訓練 (Pre-training) 以改善字向量的品質，及在生成階段設計更優良的評分方式，以利模型選擇更合理的摘要內容；另外，我們也正在研究使用卷積類神經網路 (Convolution Neural Network) 來處理篇幅更長的文章，以期建立一個更泛用的自動重寫式摘要系統。

參考文獻

[1] S.-H. Lin and B. Chen, *A Survey on Speech Summarization Techniques,* The Association for Computational Linguistics and Chinese Language Processing Newsletter, Vol. 21, No. 1, pp. 4–16, 2010

[2] I. Mani and M.-T. Maybury, *Advances in Automatic Text Summarization,* Cambridge: MIT Press, 1999

[3] C.-D. Paice, *Constructing Literature Abstracts by Computer Techniques and Prospects,* Journal of Information Processing and Management, Vol. 26, No. 1, pp. 171–186, 1990

[4] M. Witbrock and V. Mittal, *Ultra Summarization: a Statistical Approach to Generating Highly Condensed Non-extractive Summaries,* Proceedings of the 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 315–316, 1999

[5] A. Nenkova and K. McKeown, *Automatic Summarization*, Foundation and Trends in Information Retrieval, Vol. 5, No. 2–3, pp. 103–233, 2011

[6] Y. LeCun, Y. Bengio and G. E. Hinton, *Deep Learning*, Nature, Vol. 521, pp 436–444,

2015

[7] R. Sarikaya, G. E. Hinton and A. Deoras, Application of Deep Belief Networks for Natural Language Understanding, IEEE Transactions on Audio, Speech and Language Processing, Vol. 22, No. 4, pp. 778–784, 2014

[8] D. Amode et al., *Deep Speech 2: End-to-End Speech Recognition in English and Mandarin*, Proceedings of the 33rd International Conference on Machine Learning, 2016

[9] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Proceedings of Advances in Neural Information Processing Systems, pp. 1–9, 2012

[10] D. Bahdanau, K. Cho and Y. Bengio, *Neural Machine Translation by Jointly Learning to Align and Translate*, CoRR, abs/1409.0473, 2014

[11] I. Sutskever, O. Vinyals and Q. V. Le, *Sequence to Sequence Learning with Neural Networks*, Proceedings of Advances in Neural Information Processing Systems 27, pp. 3104–3112, 2014

[12] B. Hu, Q. Chen and F. Zhu, *LCSTS: A Large Scale Chinese Short Text Summarization Dataset*, Proceedings of Empirical Method in Natural Language Processing (EMNLP), pp.1967–1972, 2015

[13] A. M. Rush, S. Chopra and J. Weston, *A Neural Attention Model for Abstractive Sentence Summarization*, Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 379–389, 2015

[14] R. Nallapati, B. Zhou, C. dos Santos, C. Gulçehre and B. Xiang, *Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond*, Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL), pp. 280–290, 2016

[15] Q. Chen, X. Zhu, Z. Ling, S. Wei and H. Jiang, *Distraction-Based Neural Networks for Modeling Documents*, Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), pp. 2754–2760, 2016

[16] Y. Gong and X. Liu, *Generic Text Summarization using Relevance Measure and Latent Semantic Analysis,* Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 19–25, 2001

[17] Y.-T. Chen, B. Chen and H.-M. Wang, *A Probabilistic Generative Framework for Extractive Broadcast News Speech Summarization,* IEEE Transactions on Audio, Speech and Language Processing, Vol. 17, No. 1, pp. 95–106, 2009

[18] R. Mihalcea and P. Tarau, *TextRank Bringing Order into Texts,* Proceedings of Empirical Method in Natural Language Processing (EMNLP), pp. 404–411, 2004

[19] X. Wan and J. Yang, *Multi-document Summarization using Cluster-based Link Analysis,* Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 299–306, 2008

[20] S.-H. Lin and B. Chen, *Improved Speech Summarization with Multiple-hypothesis Representations and Kullback-Leibler Divergence Measures,* Proceeding of the 10th

Annual Conference of the International Speech Communication Association (Interspeech), pp. 1847–1850, 2009

[21] J. Kupiec, *A Trainable Document Summarizer,* Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 68–73, 1995

[22] G. Murray, S. Renals and J. Carletta, *Extractive Summarization of Meeting Recordings,* Proceedings of the 6th Annual Conference of the International Speech Communication Association (Interspeech), pp. 593–596, 2005

[23] J.-M. Conroy and D.-P. O'Leary, *Text Summarization via Hidden Markov Models,* Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 406–407, 2001

[24] D. Shen, J.-T. Sun, H. Li, Q. Yang and Z. Chen, *Document Summarization using Conditional Random Fields,* Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), pp. 2862–2867, 2007

[25] C.-Y. Lin, *ROUGE: Recall-oriented Understudy for Gisting Evaluation.* 2003 [Online]. Available: http://haydn.isi.edu/ROUGE/.

[26] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk and Y. Bengio, *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. CoRR, abs/1406.1078, 2014

[27] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, Vol. 9, no. 8, pp. 1735–1780, Nov. 15 1997. doi: 10.1162/neco.1997.9.8.1735

[28] J. L. Elman, *Finding Structure in Time*, Cognitive Science, Vol. 14, Issue 2, pp. 179–211, 1990

[29] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, *Distributed Representations of Words and Phrases and Their Compositionality*. Proceedings of the Advances in neural information processing systems 26 (NIPS 26). pp. 3111–3119, 2013.

[30] D. Bahdanau, K. Cho, Y. Bengio, *Neural Machine Translation by Jointly Learning to Align and Translate*. Proceedings of the ICLR, 2015

[31] M.T. Luong, H. Pham and C.D. Manning, *Effective Approaches to Attention-based Neural Machine Translation*, Proceedings of Empirical Method in Natural Language Processing (EMNLP), pp.1412–1421, 2015