應用變形器模型於文言文改寫白話文之研究

Using the Transformer Model for Classical to Modern Chinese Rewriting

作者：蔣珈文
單位：淡江大學資訊管理學系碩士班
Email：chiawen875@gmail.com

摘要

隨著歷史的變遷與發展，中文書面語標準與風格經歷數次大幅變化，造成現代人的文言文理解能力較弱。為了減少現代人文言文與白話文之間的理解偏差，協助理解文言文的寫作風格，本文將建立文言文生成白話文的文本改寫模型。

作法為基於文言文與白話文的平行語料，訓練近年流行的變形器模型，提取兩者關聯性，進行文言文轉換成白話文的改寫實驗。最後，以雙語評估研究BLEU指標對生成的白話文句子做評估，並對深度學習用於協助文言文理解提供個案考察報告。

關鍵詞：自然語言處理，神經機器翻譯，深度學習

1. 緒論

1.1 研究背景

中文書面資料一直都有文言文與白話文的分別，白話文起初只用於通俗文學作品，直到近代以後才在全社會上普遍應用。文言文是以先秦口語為基礎，以秦漢的經典著作為範式，並且伴隨封建主流意識形態的強化而形成的一種書面文體。在閱讀與寫作上，因為文化的變遷與發展，書面語的標準與風格有過數次大幅變化，其中文言文與白話文的風格差異最為鮮明，並未在日常生活中使用文言文閱讀寫作的人，都會在文言文理解能力上較為缺乏。[8]

近年來已經有許多研究藉由深度學習來探討自然語言處理，並且實作各種跟語言研究的模型，只要有大量語料資料輸入類神經網路進行反覆訓練就能夠生成有助於研究的成果。除了用於自然語言處理研究方面的用途以外，也能在一般的使用用途上有所幫助。

1.2 研究動機

漢語的複雜性影響我使用深度學習進行自然語言處理的相關研究。「語感」

的定義通常會理解為一種直接對語言的感覺，如果以深度學習的方式對極大量的語料資料反覆訓練，就能讓機器具有接近語感的思維方式，學到文言文的寫作風格。

本研究使用同一時代背景的中文語料，因為同一時代背景的作者其風格或思維邏輯會有一定程度的相似，例如：特定象徵的詞彙、某種觀念的傾向等，都能夠用來推測歸納當時作者在文章上的風格特徵。

本研究亦將透過機器翻譯模型呈現翻譯的效果，觀察其句子翻譯的結果，將應用模型如日常生活中所見。

1.3 研究目的

為了能有效利用機器翻譯減少文言文與白話文間理解的偏移，本研究嘗試以兩者間的對應關係進行翻譯。其中以神經機器翻譯（Neural Machine Translation）技術，結合相應的平行語料，訓練文言文及白話文基於變形器（transformer）的翻譯模型，實現文言文及白話文的翻譯。

本研究使用基於變形器的翻譯模型進行比較，觀察不同訓練集與測試集組合的翻譯成果，所使用的語料是「史記」、「漢書」以及「後漢書」[7]文本作為平行語料的來源。

1.4 論文架構

本論文共分為五個章節，前述的第一章為緒論，說明本研究的研究背景、動機與目的；第二章是文獻探討，說明研究的脈絡，論述相關的各項概念在發展過程中的變化與造成的影響；第三章為方法介紹，定義所關注的問題，說明所使用的系統架構，並詳述使用的方法；第四章說明實驗環境、實驗設計與實驗結果，並且對結果進行討論；第五章為結語與未來發展，總結本研究的實際成果，敘述研究過程上的限制，並提及未來可能的相關延伸研究方向。

2. 文獻探討

2.1 自然語言處理

本研究中關注文言文與白話文的對應關係以及將生成模型應用於翻譯的成果，而這些研究方向都屬於自然語言處理的範疇，為了達成這些目標需要這個領域的多項成果。自然語言處理要讓電腦理解人類的語言，以中文來說，需要先斷詞並理解詞，接著分析句子的語法語義，由電腦利用機器學習從訓練的資料中自動歸納得到 「規則」，並以參數的方式儲存在模型當中。

在自然語言處理的領域之中，要介紹語言模型，通常也會同時介紹 N 字詞(N-gram)。N 字詞模型是基於(n−1)階馬爾可夫鏈的一種概率語言模型，通過 n 個語詞出現的概率來推斷語句的結構。

2.2 文言文在自然語言處理的研究歷程

本研究中關注文言文與白話文的對應關係以及將生成模型應用於翻譯的成果，而文言文在自然語言處理的研究，在 2011 年，李昀燕等人發表[10]建立明清章回小說分詞和命名實體標註準則，對紅樓夢等名著採用機器輔助結合人工

分詞和標註、並進行人工校對的方式進行處理。借鑿北京大學和台灣中央研究院分詞與標註標準，其中在分詞上盡量切分到最小語意單位；命名實體標註方面有實體名與小說塑造名、虛構名交叉使用、用詞簡潔、單詞頻率高等特點，盡量建立統一的規則。在 2015 年，龐禎軍等人發表[11]提出基於對照表以及語義相關性的簡繁漢字轉換方法，主要研究對傳統方法的改良，傳統方法中只考慮一對一直接轉換，對一對多簡體字轉繁體字的考慮不夠全面導致無法有效轉換，因此提出基於對照表和語意相關性的方法，需要得到有完善規則、強區分性的詞組，隨著這部分的研究愈發完善，將可以更加圓滿的解決其中的瓶頸問題。最後於 2021 年，項潔等人提出[12]現在的歷史研究與三十年前相比有巨大的變化，大量的電子文本會對研究方法造成影響，為了應用大量的電子文獻以及數位工具，詞彙標記是其中最重要的步驟，其中有許多理論與技術上的問題需要克服。

2.3 機器翻譯

機器翻譯屬於計算語言學的領域，其研究藉由電腦程式將寫出的文字或著說出的話從一種自然語言翻譯成另一種自然語言。主要的目的是使用語言模型找出語言中的規則、趨勢或著一種特徵，如果選擇使用語料庫的技術，就能達成更加複雜的自動翻譯，包含處理不同的文法結構、詞語的辨識、慣用語的對應等。

本研究中將機器翻譯的發展歷程分為三個階段：基於規則的機器翻譯、統計機器翻譯、神經機器翻譯。

機器翻譯的概念最早於 1629 年開始發展，哲學家笛卡兒提出了世界語言的概念，意思是將不同語言卻有相同含義的詞彙以統一符號表示。直到近代因為機器的輔助，讓機器翻譯的可行性大幅提高。20 世紀初期便有多位科學家與發明家陸續提出關於機器翻譯的理論與實作的各種計畫或想法。於 Reynolds [1]記載喬治城–IBM 實驗（Georgetown–IBM experiment）。是在 1954 年 1 月 7 日由喬治城大學和 IBM 進行的實驗，成功將約 60 句的俄文自動翻譯成英文，這場實驗被視為機器翻譯可行的開端。但是 60 個來自政治、法律、數學、科學領域的句子其實都已經提前進行羅馬化，並經過精心挑選，翻譯結果才會比較成功，在當時獲得廣泛的報導。美國政府因此也開始加大資金投入計算語言學的領域進行研究。但實際上機器翻譯並沒有這麼簡單，在 1966 年自動語言處理顧問委員會（ALPAC）提出了報告後，政府投資即大幅減少。

目前機器翻譯主流的實現方式有許多類別，例如：規則法、知識翻譯、範例法、統計法、混合法、神經式等，其中有部分類別還能依實作方式進行更進一步的劃分。以下將針對被主流用於實作的機器翻譯類型作進一步介紹。

2.3.1 基於規則的機器翻譯

規則式翻譯是語言學家利用語言學知識訂定出單詞切分規則、標註規則、句法分析規則、語義分析規則、結構轉換規則等，再與目標語言的語法結構對應，置換相應詞彙，最後生成譯文，是機器翻譯的始祖。其中可以再細分為各

種子類別：轉化法（transfer-based）、中間語法（interlingual）、辭典法（dictionary-based）。

2.3.2 統計機器翻譯

統計機器翻譯採用了基於短語的機器翻譯（Phrase-Based Machine Translation），將要翻譯的句子切成數個短語，翻譯這些短語後，再重新組合成句子。這類型的機器翻譯利用了大量的平行語料（來源語言和目標語言之間的對照譯文）作為訓練資料進行統計分析，構建出統計翻譯的模型，進而使用此模型進行翻譯。統計機器翻譯的首要任務是為語言的產生構造某種合理的統計模型，並在此統計模型基礎上，定義要估計的模型參數，並設計參數估計算法。早期的基於詞的統計機器翻譯採用的是噪聲信道模型，採用最大似然準則進行無監督訓練，而近年來常用的基於短語的統計機器翻譯則採用區分性訓練方法，一般來說需要參考語料進行有監督訓練。貝氏模型（Bayesian Model）也是一種機器翻譯方法。

2.3.3 神經機器翻譯

神經機器翻譯即代表使用類神經網路來做機器翻譯。其核心概念是人類希望用計算模型再現生物的學習系統。直接模擬譯者的思路，讓翻譯成果更為自然、通順。比起過往的作法，AI 能明白一段文本的意義，而不是用查表的方式進行翻譯。

類神經網路，又稱人工神經網路是受到生物學上的神經元研究獲得啟發的數學模型，模仿生物神經網路結構和功能的，神經系統是由多個神經元所組成，彼此間透過突觸以電流傳遞訊號。設定每一個神經元為激勵函數來模擬神經細胞的行為，等於一個公式，當對神經元入一個輸入值後，經過激勵函數的運算、輸出輸出值，這個輸出值會再傳入下一個神經元。

2.4 變形器

變形器是 Google 在[3]提出的語言模型，目的是解決兩個循環神經網路（RNN，Recurrent Neural Network）系列模型遇到的問題：會遺忘早期輸入的序列，編碼結果通常無法反映完整的序列內容；輸入與輸出階段都必須依次進行，導致速度較慢。

Seq2Seq 模型是用於處理序列到序列問題，例如機器翻譯的來源語言和目的語言的句子通常沒有相同的長度，而 Seq2Seq 可以將一個序列轉化成一個不定長的序列輸出，在長短期記憶模型（LSTM，Long Short-Term Memory）作為 Seq2seq 任務模型的時候，就已經引入注意力（Attention）機制解決「模型會遺忘早期輸入的序列，編碼結果通常無法反映完整的序列內容」的問題，後來變形器模型用全注意力的結構以及進行平行運算解決「輸入與輸出階段都必須依次進行，速度緩慢」的問題，讓運算速度極大提升，基於變換器的雙向編碼器表示技術（BERT，Bidirectional Encoder Representations from Transformers）等包含巨量參數的預訓練模型也成為可能。預訓練語言模型可以從大量無標注文本中學到潛在性的語義資訊，而不需要為每一項下游自然語言處理任務單獨標

注大量訓練資料，2018 年提出的模型包括基於 RNN 的 ELMo[5]和 ULMFiT[4]，基於變形器的 OpenAI GPT（生成型已訓練變換模型，Generative Pre-trained Transformer）[6]及 Google BERT[3]等。

文字分類模型微調（Model Fine-tuning for Text Classification）是一種有效的遷移學習方法，在大規模資料集上進行預訓練（pre-training）然後在特定資料集上達到技術發展現狀（state-of-the-art），可以適用於任何自然語言處理任務，利用 Wikitext-103 語料預訓練語言模型（Language Model）並在文本情感分析（sentiment analysis）等 6 個文本分類任務上將錯誤率減少 18 至 24 個點，同時在小規模資料上也獲得較好的訓練結果（精度堪比 100 倍的資料集），另外在較小的資料集上與其他預訓練模型比起來未發生過擬合的問題。

2019 年 2 月，OpenAI 釋出 GPT-2，一個大規模無監督自然語言處理模型，可以生成連貫的文字段落，重新整理了 7 大資料集當代最佳表現，並且能在未經預訓練的情況下，完成閱讀理解、問答、機器翻譯等多項不同的語言建模任務。OpenAI 研究人員也許堅持著：「監督學習會造成語言模型僅能處理特定任務表現很好，而在泛化能力表現很差；而單純依靠訓練樣本的增加，很難有效實現任務擴充套件。」因此，他們選擇在更通用的資料集基礎上使用自注意力模組遷移學習，構建在零訓練（zero-shot）情況下能夠執行多項不同自然語言處理任務的模型。GPT-2 把變形器模型引數擴容到 48 層，包含 15 億引數的變形器模型，找到一個 800 萬網頁（WebText）資料集作為無監督訓練資料。直接以超出 10 倍的資料量上進行訓練，引數量也多出了 10 倍。這讓 GPT-2 採用更加直接的方式，單靠提升模型引數容量和增加訓練資料的數量來超過 BERT。

3. 研究方法

3.1 問題定義

中文有許多深奧的地方，發展的時間非常長，但是日常生活中已經不會使用文言文進行寫作，因此雖然是地球上重要的文化資產之一，仍然有很多人對文言文缺乏理解。而現在有大量歐美語系的翻譯模型、歐美語翻漢語也有不少，雙語種的翻譯已十分常見，文言文與白話文的翻譯卻十分缺乏。

為了能有效利用機器翻譯減少文言文與白話文間理解的偏移，本研究嘗試以兩者間的對應關係進行翻譯。其中以神經機器翻譯技術，結合相應的平行語料，訓練文言文及白話文基於變形器翻譯模型，實現文言文及白話文的翻譯。

3.2 資料前處理

資料前處理，是在將正式資料投入使用前，對資料進行整理的過程，以提升 Model 的預測效果。本研究中在訓練語料之前，要先將文言文及白話文的資料進行整理。包含：

- 資料清理：先大致檢視所有資料，並對於資料中所產生的一些錯誤或著遺漏進行清理，以提升模型的效果。
- 繁簡轉換：在字體上的繁簡轉換，因為下載的資料集是簡體字，目前大多

數系統也使用簡體字，因此我將原文保持簡體、譯文轉換為繁體字，即是使用深度學習模型進行翻譯的同時將簡體字轉繁體字。

- 再次資料清洗：對完成以上步驟的資料再做一次資料清洗，避免有可能製造的錯誤。

3.3 變形器

變形器是一個 Seq2Seq 模型，由《Attention is All You Need》[2]提出，有一個完整的編碼器-解碼器（Encoder-Decoder）框架, 其主要由注意力機制構成。模型分為編碼器和解碼器兩部分，編碼器部分由 6 個相同編碼器疊在一起，解碼器部分也是由 6 個相同解碼器疊在一起，編碼器之間不共享引數。

利用此技術訓練的任務，工程師不再需要使用 RNN+注意力機制的作法，只要專注在注意力本身就好，運算時間也相較 RNN 來得短。除了文字翻譯，該模型架構也可應用到許多自然語言處理的任務，因此衍生出許多論文及知名模型，如 BERT、GPT、XLNet. . . 。

變形器是首個完全依靠自注意力來計算其輸入和輸出表示，而不使用序列對齊的循環神經網絡或卷積的轉換模型。

3.4 BLEU 計算法

雙語評估研究指標 (BLEU，Bilingual Evaluation Understudy)是一種評估文字相似度的計算公式，是用於評估模型生成的句子(candidate)和實際句子(reference)的差異的指標。BLEU 分數取值範圍 0～1，分數越接近 1，說明翻譯的質量越高。相似度可以依據使用者的需求採用單一詞彙（1-gram）或 2-4 個字詞長度（2-4 gram），在量測相似度時不失去詞語間的順序，本研究中採用 BLEU-4，考慮 1 字詞到 4 字詞的匹配程度。

以下是 BLEU 的完整公式：

$$BLEU = BP \times exp(\sum_{n=1}^{N} W_n \times \log P_n) \tag{1}$$

$$BP = \begin{cases} 1 & lc > lr \\ exp(1 - lr/lc) & lc \leq lr \end{cases} \tag{2}$$

以上公式，BLEU 需要計算譯文 1-gram、2-gram 等，N-gram 的相似度，公式中的 P_n 指 n-gram 的相似度；W_n指 n-gram 的權重，一般設為均勻權重，即對於任意 n 都有 $W_n = 1/N$，N的上限為 4；BP是懲罰因子，如果譯文的長度小於最短的參考譯文，則BP小於 1；lc機器譯文的長度，lr是最短的參考翻譯句子的長度；1-gram 相似度表示譯文忠於原文的程度，而其他 n-gram 表示翻譯的流暢程度。

在本研究中將以 BLEU 作為評估機器翻譯品質的指標，對本研究通過模型生成的白話文句子進行評分，並根據一般情況下的範例呈現 BLEU 數值與精確度分數。

4. 實驗設計與結果

4.1 實驗環境

以下將對本研究中使用的實驗環境作介紹，包括執行環境、所使用套件和其中所使用的語料來源和內容做敘述。

本研究執行時的實驗環境，具體列表如下所示

表一：執行時的實驗環境

	作業系統	CPU	GPU	RAM	Python
執行環境	Ubuntu 18.04.1 LTS Kernel 4.15.0-47-generic #50-Ubuntu SMP	Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz	GeForce RTX 2080 Ti x2	64GB	3.7.3
	Anaconda	Keras	TensorFlow		
	4.8.3	2.7.0	2.7.0		

4.1.1 使用套件

本研究所使用到的套件，其中用於實作模型與評估功能的套件整理如下。

表二：使用套件

套件	OpenCC	Transformer[9]	numpy	NLTK
實作功能	繁簡轉換	改寫模型	張量運算	BLEU 評估指標

4.1.2 語料來源

本研究所使用到的語料為網上下載擷取的，分為文言文語料及翻譯的白話文語料，包括筆數與來源，並整理如下。

表三：語料資料

來源：https://github.com/NiuTrans/Classical-Modern			
文言文	史記	漢書	後漢書
	17701 筆	37622 筆	17753 筆
	289460 字	676569 字	311324 字
	最長 121 字	最長 320 字	最長 192 字
	最短 2 字	最短 2 字	最短 2 字
	平均句長 20.499 字	平均句長 22.658 字	平均句長 22.107 字
白話文	史記翻譯	漢書翻譯	後漢書翻譯
	17701 筆	37622 筆	17753 筆
	494470 字	1077064 字	440768 字
	最長 215 字	最長 505 字	最長 221 字
	最短 4 字	最短 3 字	最短 3 字
	平均句長 32.337 字	平均句長 33.418 字	平均句長 29.527 字

4.2 實驗設計

本研究中使用已開源的變形器模型，以網路上下載的翻譯語料進行訓練，並觀察其不同條件下的翻譯效果，包含：是否訓練過等，以及兩個模型在同一條件下的翻譯效果比較。最後通過以 BLEU 算法來對以上兩個模型生成的結果評估表現。

4.2.1 文本改寫實驗

在本研究中為了對照文言文改寫成白話文實驗的數據表現，所以無論是文言文改寫白話文或是白話文改寫文言文，限定每項評測都要使用相同的參數進行。其中 Epoch 表示將資料投入一代完整的訓練的次數，並且在每一代實驗中都不會額外進行參數調整以及除變數以外的人工調整，而是直接進行下一個實

驗。

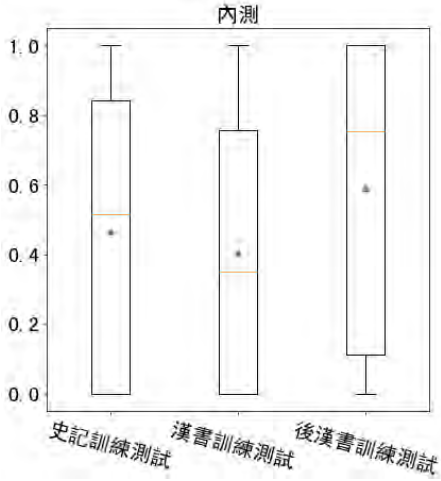
4.3 實驗結果

以下將對生成的結果以 BLEU 演算法進行整體的評估，評測的結果將以客觀的分數呈現，並且列出部分生成的結果進行直接的檢視。

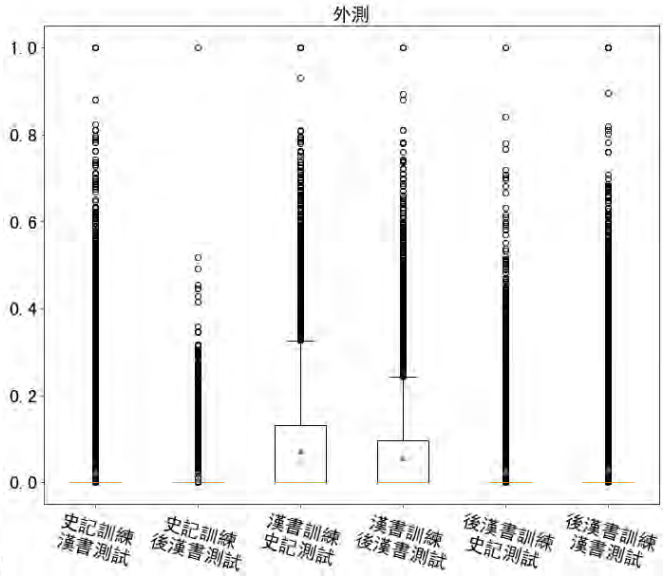
4.3.1 文本改寫結果

本研究使用機器翻譯最常使用評測結果表現的 BLEU 算法評測在文言文改寫成白話文神經機器翻譯的表現並製成表格，呈現實驗數值，並且分別列出測試集在是否參照訓練集的情況下的表現加以評估。

下方圖一、圖二中顯示文言文到白話文神經機器翻譯，圖一為使用史記訓練測試、漢書訓練測試、後漢書訓練測試的 BLEU 分數箱型圖；圖二為史記訓練漢書測試與後漢書測試、漢書訓練史記測試與後漢書測試、後漢書訓練史記測試與漢書測試的 BLEU 分數箱型圖。兩圖說明：縱軸為以資料集測試的 BLEU 分數，表示生成語句結果的相似度。

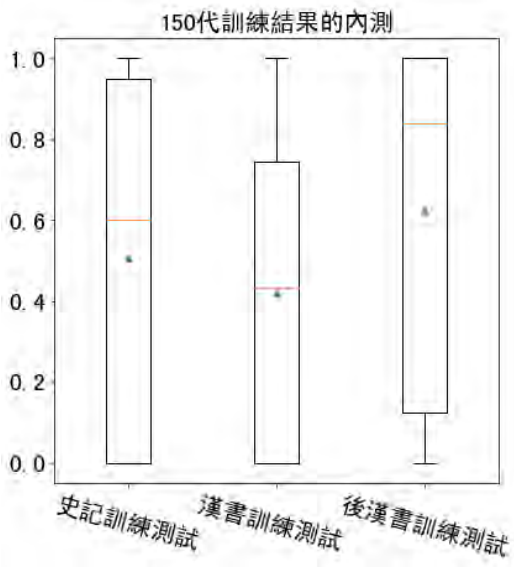


圖一：以內測模型測試結果

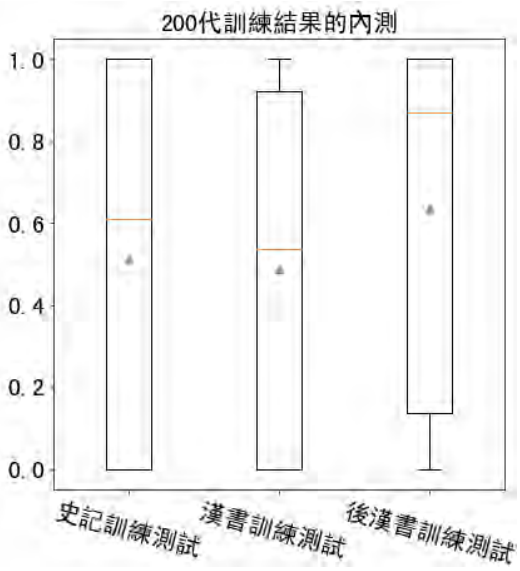


圖二：以外測的模型測試結果

下方圖三為 150 代、圖四為 200 代訓練結果的內測，圖一與圖二為 100 代，因為 100 代時在訓練時的時間長度為 1 小時上下，所以決定以 100 代為標準，接著測試代數更多的內測效果如何？結果如下所示，雖然效果有提升卻未能做到提高 0.1 的 bleu 分數，效果有限。如果能使用更多層數的變形器模型，會有更好的效果，但是需要更長的訓練時長。



圖三：訓練 150 代的內測模型測試結果



圖四：訓練 200 代的內測模型測試結果

下表為神經機器翻譯在訓練集與測試集上取得的模型平均表現分數，由以上圖一圖二可以明白 BLEU 分數的分布並不穩定，而使用與訓練集相同的測試集測試 BLEU 分數的分布會較穩定；外測以漢書訓練的模型會更穩定一些。

表四：神經機器翻譯訓練集與測試集 BLEU 評測平均得分表

文言文到白話文		
變形器模型	使用的資料集	BLEU 平均分數
使用史記訓練的模型	以史記資料集測試	0.466
	以漢書資料集測試	0.024
	以後漢書資料集測試	0.007
使用漢書訓練的模型	以史記資料集測試	0.073
	以漢書資料集測試	0.405
	以後漢書資料集測試	0.057
使用後漢書訓練的模型	以史記資料集測試	0.026
	以漢書資料集測試	0.032
	以後漢書資料集測試	0.591

以下為文言文改寫白話文神經機器翻譯中最接近 BLEU 平均值的測試集與生成結果

- 史記訓練史記測試
  - BLEU 分數：0.466
  - 測試句：廉頗遂奔魏之大梁。
  - 生成句：廉頗於是，魏國奔齊國的大梁。
  - 答案句：廉頗於是也逃奔魏國的大梁。
- 史記訓練漢書測試
  - BLEU 分數：0.024
  - 測試句：成帝初即位，丞相衡、御史大夫谭奏言：帝王之事莫大乎承天之序，承天之序莫重于郊祀，故圣王尽心极虑以建其制。
  - 生成句：成帝剛即位，對丞相的時候，對宦官十分幹中說：這是最大的不行的。

- 答案句：成帝剛即帝位，丞相匡衡、御史大夫張譚上奏說道：帝王的事沒有比繼承上天的序位更重大的了，而繼承上天的序位沒有比郊祀更重要的，所以聖王都盡心竭慮來建立郊祀的禮制。

- 史記訓練後漢書測試
  - BLEU 分數：0.012
  - 測試句：时，辽东鲜卑六千余骑亦寇辽东玄菟，乌桓校尉耿晔发缘边诸郡兵及乌桓率众王出塞击之，斩首数百级，大获其生口、牛、马、什物，鲜卑乃率种众三万人诣辽东乞降。
  - 生成句：當時，遼東郡率領軍隊伍尚出兵，兇猛獲了敵人，逃跑到曲沃郡。
  - 答案句：當時遼東的鮮卑有六千多騎兵也入侵遼東和玄菟，烏桓校尉耿曄徵調沿邊各郡的部隊和烏桓率衆王出邊塞進攻鮮卑，殺了幾百人，獲得大量的人口、牛、馬和各種器物。鮮卑首領這才率領三萬部衆到遼東乞求投降。
- 漢書訓練史記測試
  - BLEU 分數：0.073
  - 測試句：赐陆生橐中装直千金，他送亦千金。
  - 生成句：賜陸賈的珠寶價值千金，別贈送其別十千金。
  - 答案句：尉他還送給陸生一個袋裝包裹，價值千金，另外還送給他不少其它禮品，也價值千金。
- 漢書訓練漢書測試
  - BLEU 分數：0.405
  - 測試句：当是时，丞相入奏事，语移日，所言皆听。
  - 生成句：那個時候，丞相入內奏事，往一坐移到一天，他所說的是要排列的情況。
  - 答案句：在那個時候，丞相入內奏事，往往一坐就是大半天，他所提的意見皇上一概接受。
- 漢書訓練後漢書測試
  - BLEU 分數：0.057
  - 測試句：弼大怒曰：太守忝荷重任，当选士报国，尔何人而伪诈无状命左右引出，楚捶数百，府丞、掾史十余人皆谏人廷，弼不对。
  - 生成句：她們大怒說：太守、太守的人都是不高興的，應當挑選人談，這些近臣性情不知。
  - 答案句：史弼大怒說：太守身當重任，應當選士報國，你是什麼人，這樣欺詐無體統命令左右的人把他帶出去，打了幾百大板，府丞、掾史十幾個人都在公廷勸諫他，史弼不理。
- 後漢書訓練史記測試
  - BLEU 分數：0.026

- 測試句：李敢以校尉从骠骑将军击胡左贤王，力战，夺左贤王鼓旗，斩首多，赐爵关内侯，食邑二百户，代广为郎中令。
- 生成句：鄧成都尉驃騎將軍左右的左昌，時力有申騎，後左右、受城郎，斬宗族都尉。
- 答案句：李敢以校尉官職隨從驃騎將軍出擊匈奴左賢王，奮力作戰，奪得左賢王的戰鼓和軍旗，斬殺很多敵人首級，因而賜封了關內侯的爵位，封給食邑二百戶，接替李廣任郎中令。

- 後漢書訓練漢書測試
  - BLEU 分數：0.032
  - 測試句：平帝元始三年正月，天雨草，状如永光时，京房《易传》曰：君吝于禄，信衰贤去，厥妖天雨草。
  - 生成句：平帝大元年正月，天下，老雨希望着永，岱宗至美，上天德傳，彰美陽。
  - 答案句：平帝元始三年正月，天空落下草來，樣子和永光二年那次一樣。京房《易傳》說：國君少給了俸祿，忠信就要消失，賢人就要離去，這時出現的妖祥就是天上下草。
- 後漢書訓練後漢書測試
  - BLEU 分數：0.591
  - 測試句：彭宠为太守，召延署营尉，行护军。
  - 生成句：彭寵做太守，召蓋延署職營尉，代理護軍，代他的軍赴任。
  - 答案句：彭寵爲太守，召蓋延署職營尉，代理護軍。

5. 結語與未來發展

5.1 結論

為了盡可能消除理解文言文的偏差，本研究將使用兩者的平行語料作為訓練集來提取兩者間的對應關係，並以變形器模型生成對應語句以實現文言文改寫白話文的目的。

對以上目標，本研究以 BLEU 指標進行評估，取得客觀分數標準並互相對照，以認知實際上系統的表現。因此，本研究完成如下成果：

一、比較史記、漢書、後漢書三種資料集在個別建立文言文改寫白話文模型的內測 BLEU 表現分數，其中後漢書訓練測試的結果最好，史記次之，漢書最差。

二、比較史記、漢書、後漢書三種資料集在個別建立文言文改寫白話文模型的外測 BLEU 表現分數，其中漢書訓練史記測試與後漢書測試的結果稍好，其它很差。

三、比較史記、漢書、後漢書三種資料集在個別建立、不同訓練代數的文言文改寫白話文模型的內測 BLEU 表現分數，代數越多，結果越好。

5.2 研究限制

首先，因為文言文與白話文在結構上的差異，導致本研究的一些研究困難，文言文的風格十分簡練，造成白話文的翻譯需要大段解釋其句子中的含意，因此在平行語料的詞對應上通用性較弱。

除此之外，關於本研究的實驗數據，因為研究環境的運算效能及運算速度的限制，無法完整實測各種參數條件的可能如果能持續調整語料資料、設置的參數、增加額外的語料，模型的表現應該還有提升的可能性。

5.3 未來展望

本研究使用變形器模型進行翻譯並且以 BLEU 指標評估文言文對白話文神經機器翻譯的成果，未來可以使用美國國家標準技術研究所（NIST，National Institute of Standards and Technology）評測方法、錯詞率(The Word Error Rate，WER)、METEOR 測度等方法進行機器翻譯生成結果的評估，並且與本研究的 BLEU 評估分數並列。

參考文獻

[1]、 A. C. Reynolds, J. (1954). THE CONFERENCE ON MECHANICAL TRANSLATION Held at M.I.T., June 17-20, 1952

[2]、 Ashish Vaswani, N. S., Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. (2017). Attention Is All You Need.

[3]、 Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[4]、 Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.

[5]、 Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv: 180205365. In.

[6]、 Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[7]、 古文现代文翻译平行语料库. Retrieved from https://github.com/BangBOOM/Classical-Chinese

[8]、 蔣宜衡. 應用序列到序列生成模型於雙向文本改寫之研究.淡江大學碩士論文(2018)

[9]、 淺談神經機器翻譯 & 用 Transformer 與 TensorFlow 2 英翻中 https://leemeng.tw/neural-machine-translation-with-transformer-and-tensorflow2.html

[10]、 李昀燕;熊丹;陸勤;石定栩;趙天成. 明清章回小說的分詞準則及命名

實體標註. 元智大學財務金融學系(2011)
[11]、 庞祯军，姚天昉. 基于对照表以及语义相关性之简繁汉字转换[J]. 计算机工程与应用, 2015, 51(4): 115-119.
[12]、 项洁, 胡其瑞. 历史文本的词汇标记及其应用[J]. 数字人文研究, 2021, 1(1): 48-64.