

基於深度學習之簡答題問答系統初步探討

A Preliminary Study on Deep Learning-based Short Answer Question Answering System

林鈺宸 Yu-Chen Lin, 廖元甫 Yuan-Fu Liao
國立臺北科技大學電子工程系

Department of Electronic Engineering, National Taipei University of Technology
t106368024@ntut.edu.tw, yfliao@ntut.edu.tw

Matúš Pleva, Daniel Hládek

Department of Electronics and Multimedia Communications, Technical University of Košice,
Slovakia

matus.pleva@tuke.sk, daniel.hladek@tuke.sk

摘要

一般問答系統常是以從題目原文中尋找最可能的文字段落的方式來生成答案，但是最近開始有試圖以編碼器(encoder)先抽取題目與問句的隱含意義，再以解碼器(decoder)重新生成回應語句的趨勢。因為它能不受原始文章使用的文字的限制，甚至可以用完全不同的說法來回答。因此在本論文中提出了一個基於 BERT 與 Transformer 的“編碼-解碼”模型，嘗試實現此問答系統。實驗方面使用台達電研究所的公開數據集(Delta Reading Comprehension Dataset, DRCD[4])和『科技大播臺，與 AI 對話』(Formosa Grand Challenge)的比賽資料，來訓練模型。由實驗結果發現，我們的系統可以由學習產生答案，並達到 ACC=66.22%，F1=41.16%，EM=18.41%，BLEU=3.98%的效能，嘗試實現可自由產生回應語句的問答系統。

關鍵詞：自然語言處理、問題回答、序列到序列模型, NLP, QA, Seq2Seq

一、緒論

在自然語言處理中的問題回答系統領域有許多不同類型的文本理解競賽，如科技部舉辦的『科技大播臺，與 AI 對話』(Formosa Grand Challenge) 競賽、史丹佛大學發起的挑戰(Stanford Question Answering Dataset, SQuAD[3])以及分別來自於 CMU、Stanford 和 Mila 的學生推出的新型問答挑戰 HotpotQA[5]等等。

因為問題回答系統領域在國外是一個很熱門的挑戰，所以延伸許多不同的閱讀理解數據集，如史丹佛的 SQuAD(表 1)，它不同於一般的閱讀理解數據集，在它的答案不是僅僅一個詞或是物體，而有可能是一段句子，使得答案更難以預測，但是此數據集的問題大多可以使用關鍵字匹配的方式來進行搜索回答；而 HotpotQA(表 2)的問題被設計成

需要以多步推理的方式來回答，所以無法輕易地以關鍵字匹配的方式來解答。另外，此數據集沒有預設任何的知識圖譜，因此問題內容具有多樣性，使得它更有難度。

由於受到了國外的刺激，我國政府機關也開始積極投入這個領域，所以有了 Formosa Grand Challenge 的比賽，此數據庫它需要有對問題的推理能力還有閱讀的理解能力，且不同於 SQuAD 和 HotpotQA 的閱讀理解數據集，它無法輕易以搜索答案在文章中的起始與結束位置得知正確答案，比賽的簡答題的部分需要生成一段完整的句子且需要契合參考答案。

表 1、SQuAD 閱讀理解數據集範例

C	Victoria (abbreviated as Vic) is a state in the south-east of Australia. Victoria is Australia's most densely populated state and its second-most populous state overall. Most of its population is concentrated in the area surrounding Port Phillip Bay, which includes the metropolitan area of its capital and largest city, Melbourne, which is Australia's second-largest city. Geographically the smallest state on the Australian mainland, Victoria is bordered by Bass Strait and Tasmania to the south,[note 1] New South Wales to the north, the Tasman Sea to the east, and South Australia to the west.
Q	Where in Australia is Victoria located?
A1	south-east
A2	the south-east of Australia

表 2、HotpotQA 閱讀理解數據集範例

C1	Gorgeous George (album) Gorgeous George is the third solo studio album by Scottish musician Edwyn Collins. The album was recorded at New River in London, with Collins acting as the producer.
⋮	⋮
C10	Jimmie Ross Jimmie Ross is an American rock guitarist and vocalist who is best known for being a member of Pittsburgh band the Jaggerz, known for their 1970 hit. During the band's initial existence of 1965-1976, the bassist shared the duties of lead vocalist with guitarist Donnie Iris. By the time the Jaggerz regrouped in 1989, Iris was well into his solo career, and Ross became the sole lead vocalist and remained bassist.", " He continues to hold both positions today.
Q	Which musician, Edwyn Collins or Jimmie Ross, played the bass guitar?
A	Jimmie Ross

表 3、Formosa Grand Challenge 閱讀理解數據集範例(簡答題)

C	請聽這段話，然後回答問題： 科學研究結果顯示，人類種植可可樹的起源可以追溯至大約 3600 年前。研究人員抽絲剝繭，分析了 200 棵可可樹的基因組之後認定，最早種植可可樹的是現今厄瓜多境內古瑪雅人。科學家也指出，早期的可可豆風味相當不同。人工種植可可樹後，人類選擇自己偏好的性狀，然後不斷繁殖，並持續改良植物的大小和風味等特徵。在培育一代又一代可可樹的過程中，可可豆
---	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	的風味逐漸改變，苦味變濃，帶給人興奮感的可可鹼含量也逐漸增加。但在此同時，可可樹對病蟲害的抵抗力也隨之下降，使得它愈來愈珍貴。
Q	為什麼可可樹的價值逐漸增高?
A1	可可樹易得病蟲害，種植成本增加，數量減少。
A2	因為可可豆的風味逐漸改變，帶給人興奮感的可可鹼含量逐漸增加，但同時可可樹對病蟲害的抵抗力也下降了。

為了解決 SQuAD 的類型的題目，目前的方式有 QANet[1]和 BERT[2]與 Transformer[9]組成的 Seq2Seq 模型，它們基於開放性領域的知識庫上建立一個問題回答系統需要透過既有的知識庫中來進行特徵提取，之後將向量將抽取特徵和相同意義的資訊建立在一起進而會答問題。而動態融合圖形網路(Dynamically Fused Graph Network for Multi-hop Reasoning, DFGN[17])是為了解決 HotpotQA 而發展出來的架構，它能從多個片段文字段落動態建構的特徵進行探索，並從特定的文章中逐步找出的相關資訊進行推理來找尋答案。

在 Formosa Grand Challenge 的比賽中我們測試使用 QANet 和 DFGN 的架構配合 DRCD[4]所訓練出來的模型進行問題的解答但其效果都有限，主要是因為這些過去的任務主要還是以搜索答案的位置為主，因此如果一個正確答案不在文章中系統便無法回答。雖然在比賽中還是會遇到單純的分析問題，但在每次比賽過程中會慢慢發現這種分析問題越來越少，且在決賽時會有簡答題的題型，為此我們認為需要一套可以了解文章內容並且可以自然答案生成的系統。

在 Formosa Grand Challenge 的比賽中需要了解文章內容進而回答問題，為了配合後續簡答題的回答策略，所以我們在選擇題的部分先以回答簡答的方式生成答案，再配合提供的選項來進行匹配。我們嘗試以編碼器先抽取題目與問句的隱含意義，再以解碼器重新生成回應語句，而不是從題目原文中尋找最可能的文字段落的方式來生成答案最後在與答案做關聯性的匹配。這不同於過去的知識庫回答系統，他們比起回答一個擁有正確答案且完整句子，更注重於分析問題和搜尋相關答案。我們實驗了(如圖 1)基於注意力機制的自然答案生成模型，在系統裡使用多組閱讀測驗問答式的資料進行訓練，並使用序列到序列(Seq2Seq)的學習框架來完成任務。

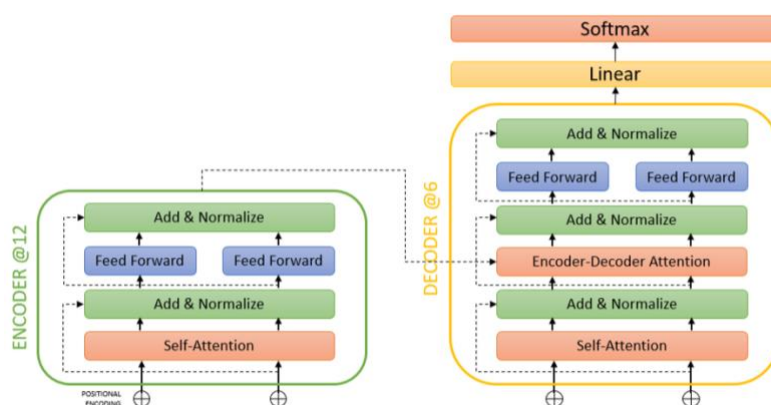


圖 1、模型架構示意圖

二、相關研究

Seq2Seq 模型架構最初提出是為了機器翻譯[12]，它解決 RNN 不定長度的問題並能有效地建立一個基於輸入序列用來預測未知輸出序列，在這之後此架構在機器翻譯、寫作、文本摘要和人機對話等主題有更多的發揮。最初所提出的 Seq2Seq[18]模型想要在一個完整的輸入序列中找出語意來做向量編碼是一件很困難的事情，特別是對於較長的輸入，因此提出了注意力機制[19]的 Seq2Seq 模型架構。由於我們用於訓練所使用的資料經常以較長的輸入為主且需要尋找出關鍵的語意進而回答問題，所以我們傾向使用擁有注意力機制的 Seq2Seq 模型架構來培訓我們的模型。

過去常用於問題回答的方法從 RNN 架構，到近幾年的 QANet 和 DFGN 等等。QANet 它拋棄了 RNN 的作法並設計了一個模型架構，由多個編碼區塊建立而成的，在每個區塊中都包含 Multi-Convolution-Layer、Self-Attention Layer[16]和 Feedforward Layer。其知識庫基本是以詞為單位，使用 200 和 300 不同維度的詞向量(Word2vec[6-8])為基礎，再透過編碼器來學習文章與問題之間的關聯性，由於使用 Self-Attention 因此對全局的資訊做有效的處理來得到較好的結果。

為了改善知識庫的部分而提出新穎的預訓練方法 BERT 並配合 Transformer 組成一個是 seq2seq 模型，BERT 的部分使用 Transformer 的架構來訓練單字級別的知識基礎，由於使用 Transformer 所以它相較於深層遞迴類神經網路(RNN)更能有效捕捉較長句子的資訊，使它在問答解題相較於 QANet 有較好的結果。

三、方法

序列到序列模型架構

過去傳統序列到序列(Seq2Seq)模型架構應用在一連串有相關的連續數據中，如語音數據、影像數據等，因為它具有輸入以及輸出長度不固定的特性以及元素之間的順序關係。在所有的問答題中它們的文章、問題以及答案的長度並不是等長的，而在輸出答案的前後順序也會影響著我們得到結果，因此我們採用此架構為我們的訓練框架。

Seq2Seq 分成兩個部分，編碼器和解碼器。編碼器我們使用 Google 的 BERT 預訓練模型，首先將文章與問題串連後透過 Tokenizer 轉換成輸入序列後再輸入到解碼器得到編碼資訊。在解碼器的部分我們使用 Transformer 的架構，總共有 6 層，它將從編碼器所得到的編碼資訊配合上一個輸出來產生解碼器的輸出資訊。最後，將輸出資訊連接到線性的全連接層透過 softmax 輸出成語字典同樣的維度。

(一) BERT 神經網路模型

BERT 重點在於它能貫穿每層的背景關係來預訓練深度的雙向表示，它與其他模型不同的地方在於它多個一個機制稱為 Masked LM (MLM)，能隨機地屏蔽部分的輸入 token 只計算被遮蓋掉的 token，目的是根據被遮掉標記的背景關係來預測原始對應的 token，另外使用雙向 Transformer，能讓 MLM 標記可以學習到前後背景關係。

在實驗中我們先將文字以既有的辭典映射成數字的形式來表示，BERT 模型輸入的

地方有兩個輸入，第一個部分是輸入文字序列(Token)，第二個是輸入分割序列(Segment)。在 Token 的部分給出一個中文單字級別的序列 $[101, q_0, \dots, q_L, 102, c_0, \dots, c_L, 102]$ ，它包含了問題以及文本的文字編碼，問題前以及文本之間和最後都以特殊字元隔開，問題 q 和文本 c 的總長度為 L 。另外，在 Segment 的部分給入一個由 0 和 1 組的序列，它是根據文章和問題來做分割的，在這邊我們將問題依照 Token 的數量給予相同的 0，而文章給予 1，因此最後會形成一個序列 $[0, \dots, 1, 1, \dots, 1]$ 用來區分文章與問題。最後，在將得到的文字序列和分割序列透過 Word Embedding 輸入進 BERT Model，得到該序列的編碼資訊。

(二) Transformer 模型架構

解碼器是使用 Transformer 架構組成的，將訓練答案以跟 BERT 神經網路模型所使用的辭典映射成數字的形式 $[101, a_0, \dots, a_L, 102]$ ，之後轉換成訓練對 $[101, a_0, \dots, a_L]$

和 $[a_0, \dots, a_L, 102]$ ，在訓練過程中將從上一個解碼器的輸出值作為輸入，來訓練下一個輸出值的機率分布。應用於解碼器的 Transformer 有一個不同，它多增加一個 Self-Attention 單元，由於序列到序列的模型在訓練時的解碼器需要參考上一個輸出的結果來當作下一個的解碼器的輸入，而編碼器又需要根據 BERT 所分享的資訊來進行運算，因此這邊會先將上一個的輸出會透過 Self-Attention 計算後再跟 BERT 的輸出合併做運算，在來導入線性的全連接層透過 softmax 輸出成語字典同樣的維度，最後將輸出的值再透過字典轉回文字供使用者閱讀。在訓練過程中我們是以字對字的方式去學習，從 input 的每一個字去學習 output 相對應的字的出現機率。

四、實驗結果

(一) 數據集

我們使用兩個數據集來評估我們的方法，第一個是台達電研究院(DRCD)的公開的數據集，另一個是和『科技大播臺，與 AI 對話』(Formosa Grand Challenge)的比賽數據集(如表 4)。

DRCD 是一個中文閱讀理解數據集，它從 2,108 篇維基百科的文章整理出 30,000 多個問題，人類表現在此數據集的 F1 成績為 93.30%以及 Exact Match 的成績為 80.43%。Formosa Grand Challenge 它是由中華民國政府部門舉辦的一個大型競賽的中文數據集，包含了閱讀理解以及單純的問題回答(問題解答不包含在文本之中)，它從網路新聞、古文以及小說整理出來的 14,000 多個問題。另外，此數據集有 25 題是問答題，由多位專家來出題，且經過討論後給定多個參考解答，並由評審委員修改。它總共分成 6 次初賽、加賽、複賽和決賽，因此我們將初賽和複賽歸納至訓練集並隨機抽出 1,000 題來做驗證並將加賽和決賽的簡答題題目歸類到測試集，在這邊我們捨棄掉大部分決賽的資料是因為決賽中大量的問題需要從答案來判斷答案是否正確。

表 4、數據集統計

資料集	訓練	驗證	測試
-----	----	----	----

DRC	26,932	3,524	3,485
Formosa Grand Challenge	11,551	1,000	1,025

由於 Formosa Grand Challenge 比賽形式以選擇題為主，因此我們將正確的選項內容當作答案文字內容來進行訓練，另外由於它的資料庫有多餘的換行以及特殊符號，所以我們也將它刪除掉，以保持訓練文本的統一性。

(二) 評估指標

在進行多類別的標籤分類中採用分類交叉熵(Categorical Cross-Entropy)來幫助我們了解訓練上的差異，其方程式如下所示：

$$loss = - \sum_{i=1}^n \hat{y}_{i1} \log y_{i1} + \hat{y}_{i2} \log y_{i2} + \dots + \hat{y}_{im} \log y_{im}$$

其中， n 是樣本數量， m 是分類數量，由於此函數是一個多輸出的 loss 函數，所以計算出來的值也是多個的。

另外我們使用了幾個指標來幫我們評估我們的方法：第一個是正確度，主要是為了知道它保存了多少資訊在裡面，其方程式如下所示：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

其中 True Positive(TP)是將正確的預測為正確，True Negative(TN) 是將錯誤的預測為錯誤，False Positive(FP) 是將錯誤的預測為正確最後 False Negative(FN) 是將正確的預測為錯誤的。

第二個是 F1 值，統計學中用來衡量分類模型精確度的一種指標。F1 值就是精確率(precision)和召回率(recall)的調和均值。精確率它計算所有正確被檢出的結果(TP)占實際上被檢索到的(TP+FP)比例，其方程式如下所示：

$$P = \frac{TP}{TP + FP}$$

而召回率它是計算所有正確被檢出的結果(TP)占所有被應該檢索到的(TP+FN)比例。

第三個是 Bilingual Evaluation understudy(BLEU)，評分關鍵在於如何定義生成答案與參考答案之間的相似度，首先須計算詞在句中的匹配數，其方程式如下所示：

$$Count_{clip}(word) = \min\{Count(word), MaxRefCount(n - gram)\}$$

其中， $Count(word)$ 表示詞在生成文中的出現次數， $MaxRefCount(word)$ 是該參考答案的最大次數。出現詞的精確度 P_n 定義為：

$$P_n = \frac{\sum_{C \in candidates} \sum_{n-gram \in C} Count_{clip}(n - gram)}{\sum_{C \in candidates} \sum_{n-gram \in C} Count(n - gram)}$$

透過上述結果來計算 BLEU，其評分方程式如下：

$$BLEU = BP * \exp \left[\left(\sum_{n=1}^N W_n \log p_n \right) \right]$$

其中BP是懲罰值，其中 W_n 表示詞出現的權重：

$$BP = \begin{cases} 1 & \text{if } c > r \\ EXP(1 - r/c) & \text{if } c \leq r \end{cases}$$

最後是 Exact Match(EM)，它能讓我們觀察生成出的答案與標準答案是否一致。

(三) 實驗設置

訓練過程中，我們將兩個資料庫的訓練集以及驗證集各自合併一起訓練，一開始先設定文本和問題的總長度限制在 384，它可以完全覆蓋數據集中 95%的資料，另外將答案的總長度限制在 30，它能完全覆蓋 98%的答案數據。使用的辭典包含中英文以及各式標籤總共有 21128 個字。

模型架構使用序列到序列模型，在編碼器使用 BERT 中文預訓練模型，它是由 12 層雙向 transformer 組成，隱藏層擁有 768 個節點，Multi-head self-attention 的 heads 為 12；解碼器則是有 6 層雙向 transformer 組成，且每層都共享了編碼器的輸出。

(四) 結果與討論

從損失值(如圖 3)和正確率(如圖 4)可以證明我們的模型架構是可以有效的學習，而正確率越高表示我們的模型在學習過程中透過由 BERT 給予的知識以及上一次的輸出能生成出包含正確的資訊越好。我們所提出的方法用於兩個數據集的測試集，在 DRCD 的正確率擁有 66%，F1 評分有 41.16%，EM 評分有 18%；Formosa Grand Challenge 的正確率有 58%，F1 評分有 37.46%，EM 評分有 16%。如表 5。

表 5、數據集指標

資料庫	Acc	F1	EM	BLEU
DRCD	66.22	41.16	18.41	3.98
Formosa Grand Challenge	58.10	37.46	16.35	3.25

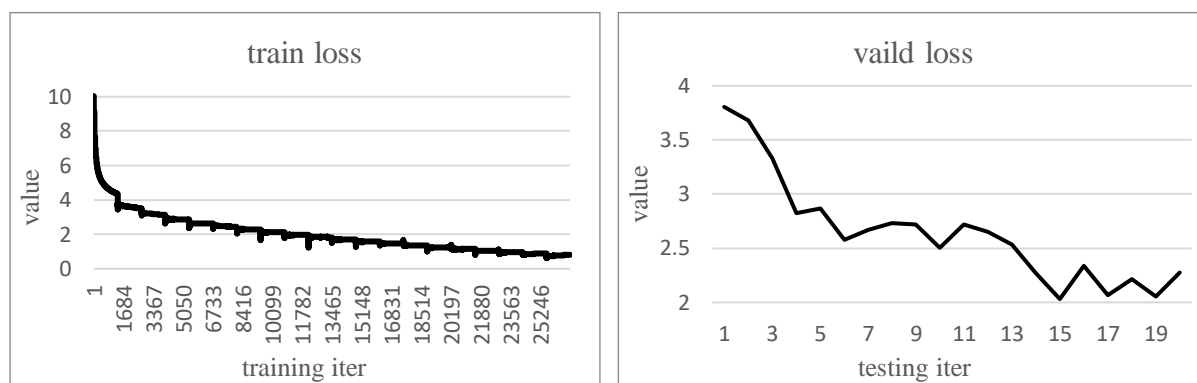


圖 3、訓練/測試的損失值分布

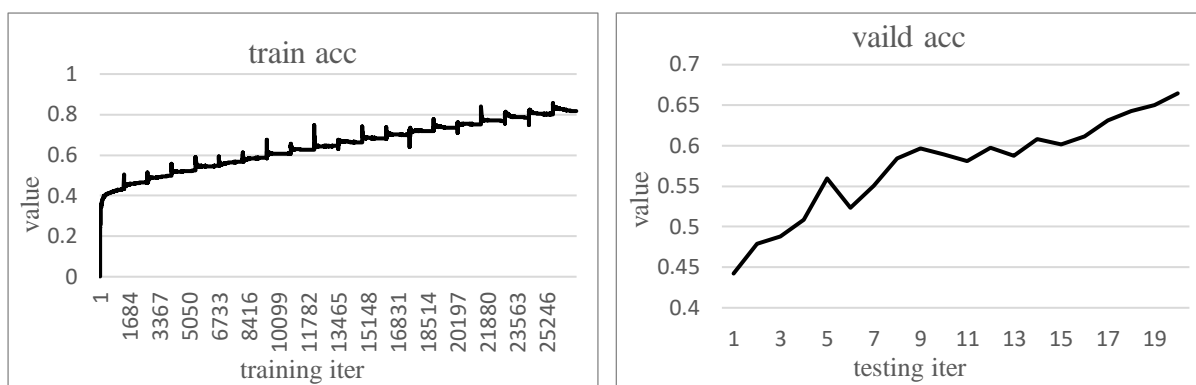


圖 4、訓練/測試的正確率分布

以下我們提出幾個案例來進行討論：從範例觀察我們發現從表 6-8 可以看到我們的模型生成流暢的句子且相似的答案，並具有正確的知識。然而答案還是存在一些問題，例如無法找出問題與文章之間正確的關鍵點，所以回答出的錯誤的答案或產生連續重複的詞彙、回答不完整，以及生成出不流暢的句子。

從上述問題中我們可以發現其實 BERT 是從一個輸入序列(文章)中找尋與問題最具有關聯性的訊息，它還不能算真正擁有解讀文章意義的模型。產生連續重複的詞彙的問題是因為序列到序列模型過度依賴上一次的輸出，因此如果出現一個重複的單字就會出現無限循環，導致產生出的答案出現重複詞彙的現象。最後生成出不流暢的句子，我們認為是訓練答案的資料的類型種類不夠多，由於我們的訓練的答案主要還是以簡短的幾個字詞或專有名詞來輸出，所以在遇到需要回答較長類型的答案時便無法生成出來。

表 6、輸出結果範例一

C1	請聽這段話，然後回答問題： 台灣宇博(Uber)公司遭公路總局罰款 3000 多萬一事，台北高等行政法院今天做出判決，判 Uber 勝訴，無須繳交罰款。根據調查，公路總局發現，Uber 未經核准，於民國 105 年間透過網路招募司機，分別在台北市、新北市等地，藉由 Uber APP 程式平台，指揮調度 233 輛車營業載送客人。載客完成後，收取報酬，違反汽車運輸業管理規則，因此處以總計逾 3000 萬元罰款。台北高等行政法院法官調查，計程車客運業的主管機關並非公路總局，而是市政府。因此撤銷處分，判 Uber 勝訴免罰。全案可上訴。
Q1	台灣宇博(Uber)公司分別在台北市、新北市兩地違法行事，應該由誰開出罰單？
A1	台北市政府、新北市政府
Our	以上皆可

表 7、輸出結果範例二

C2	請聽這段話，然後回答問題： 台東聖母醫院是個小型醫院，除了醫治病人，還肩負起照顧台東弱勢原住民的重責大任。它的廚房每天提供午餐給獨居老人和兒童，讓他們不至於挨餓。 但七月份的強烈颱風重創台東。聖母醫院的倉庫被淹沒，導致白米泡水，連供應午餐的廚房設備也嚴重毀損。醫院估計，重建經費逼近千萬。 聖母醫院表示，他們每天用的食材一向靠志工栽種的有機蔬果供應；此外，醫院向政府承
----	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	租農地，一年兩穫的稻米還可以小包裝透過網路販售。扣除成本之後，販賣的盈餘就是維持農場和廚房營運的資金。 然而風災過後，收成全無，設備毀損，復原之路十分漫長。聖母醫院因而呼籲各界人士慷慨解囊，協助早日重建，讓他們的廚房不至於斷炊。
Q2	聖母醫院平日是靠什麼樣的收入來維持廚房的營運？
A2	靠銷售自己種植的稻米所得盈餘來維持。
Our	自給農業

表 8、輸出結果範例三

C3	A：我們這週末要帶小朋友們去動物園。 B：他們喜歡什麼動物？ A：他們都喜歡大熊貓、無尾熊。哥哥特別想要看老虎和蛇，但妹妹會害怕。妹妹喜歡小兔子和鴨子，可是哥哥沒興趣。兩個吵吵鬧鬧的，真是讓人傷腦筋。
Q3	哪一種動物可以讓兩位小朋友一起看？
A3	無尾熊
Our	大熊貓

五、結論

本論文提出了一個基於 BERT 與 Transformer 的“編碼-解碼”問答模型,嘗試實現可以回答任意簡答題的問答系統,尤其是可以回答文章中沒有直接提到的答案。其能從文章中搜尋跟問題相關的資訊,並通過自我關注的機制,提高自然答案生成系統的性能。經由實驗結果驗證,我們的系統確實可以由學習產生答案,並達到 ACC=66.22%, F1=41.16%, EM=18.41%, BLEU=3.98%的效能,初步實現了可自由產生回應語句的簡答題問答系統。未來,將往兩個方向進行研究。包括(1)如何讓編碼器能擷取出更豐富的語意特徵,與(2)如何讓解碼器能對問題的理解能有更深一步的能力。

Acknowledgements

This work was partly supported by Slovak Research and Development Agency under contract no. APVV SK-TW-2017-0005, APVV-15-0517, APVV-15-0731, partly Cultural and educational grant agency from project KEGA 009TUKE-4/2019 and partly Scientific grant agency by realization of research project VEGA 1/0511/17 both financed by the Ministry of Education, Science, Research and Sport of the Slovak Republic and finally by the Taiwan Ministry of Science and Technology MOST-SRDA contract No. 107-2911-I-027-501, 108-2911-I-027-501, 107-2221-E-027-102, 107-3011-F-027-003 and 108-2221-E-027-067.

參考文獻

- [1]. Yu, Adams Wei, et al. "Qanet: Combining local convolution with global self-attention for reading comprehension." arXiv preprint arXiv:1804.09541 (2018).
- [2]. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding."

- arXiv preprint arXiv:1810.04805 (2018).
- [3]. Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).
 - [4]. Shao, Chih Chieh, et al. "Drcd: a chinese machine reading comprehension dataset." arXiv preprint arXiv:1806.00920 (2018).
 - [5]. Yang, Zhilin, et al. "Hotpotqa: A dataset for diverse, explainable multi-hop question answering." arXiv preprint arXiv:1809.09600 (2018).
 - [6]. Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).
 - [7]. Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." International conference on machine learning. 2014.
 - [8]. Bojanowski, Piotr, et al. "Enriching word vectors with subword information." Transactions of the Association for Computational Linguistics 5 (2017): 135-146.
 - [9]. Shao, Taihua, et al. "Transformer-Based Neural Network for Answer Selection in Question Answering." IEEE Access 7 (2019): 26146-26156.
 - [10]. Lin, Yuhua, and Haiying Shen. "SmartQ: A question and answer system for supplying high-quality and trustworthy answers." IEEE Transactions on Big Data 4.4 (2017): 600-613.
 - [11]. Karimi, Elaheh, Babak Majidi, and Mohammad Taghi Manzuri. "Relevant Question Answering in Community Based Networks Using Deep LSTM Neural Networks." 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS). IEEE, 2019.
 - [12]. Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. 2014.
 - [13]. Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).
 - [14]. Tu, Zhaopeng, et al. "Modeling coverage for neural machine translation." arXiv preprint arXiv:1601.04811 (2016).
 - [15]. Tan, Chuanqi, et al. "Context-aware answer sentence selection with hierarchical gated recurrent neural networks." IEEE/ACM Transactions on Audio, Speech, and Language Processing 26.3 (2017): 540-549.
 - [16]. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.
 - [17]. Xiao, Yunxuan, et al. "Dynamically Fused Graph Network for Multi-hop Reasoning." arXiv preprint arXiv:1905.06933 (2019).
 - [18]. Vinyals, Oriol, and Quoc Le. "A neural conversational model." arXiv preprint arXiv:1506.05869 (2015).
 - [19]. Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).