

第一題

A. encoder-only model:

在 Bert 的部分，需要 fine-tuning Bert 的原始模型，模型的 input 需包含問題與參考的文章(如圖 1 所示)，然後模型會將參考文章中的每個單字做兩種標記，分別是 start 跟 end，然後對每個單字做評分，最後的出一個 start 分數最高的字及 end 分數最高的字，而 start 分數最高的字就是模型最後 output 的開始位置，end 分數最高的字就是模型 output 的結束位置(如圖 2 所示)，以圖 1 的問題為例，此題的答案就會是 340 開頭 M 結尾，因此答案為 340M。

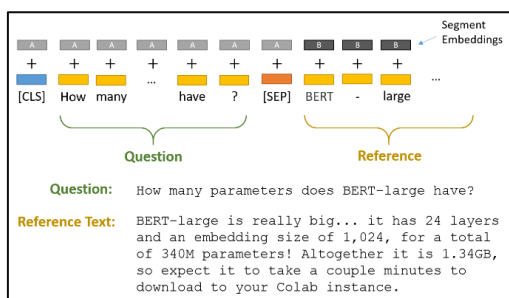


圖 1. Bert 的 input 形式

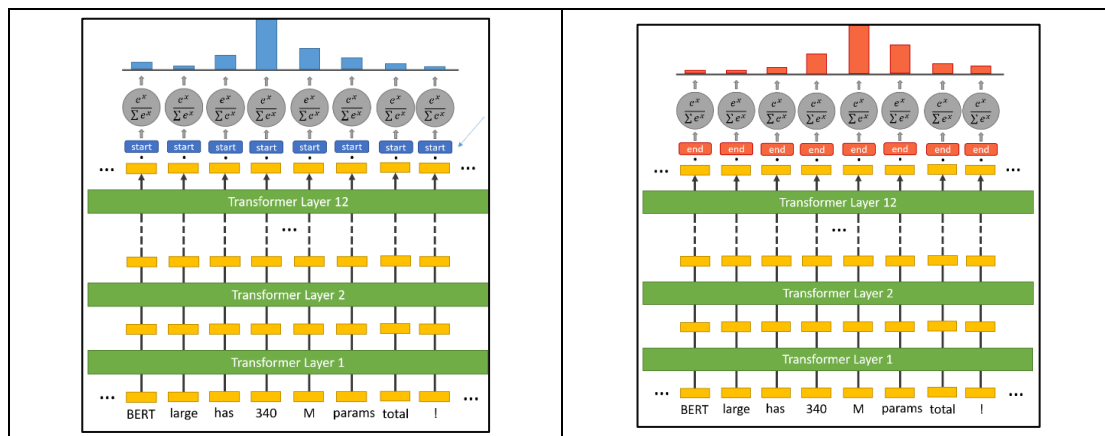


圖 2. 模型 output 的決定方式

B. decoder-only model:

在 decoder-only model 這類的模型中，模型是採用文字接龍的方式產生答案，經過學習下一個字生成什麼字的機率最大，進而回答出問題所需的答案。

參考資料:

1. <https://mccormickml.com/2020/03/10/question-answering-with-a-fine-tuned-BERT/?ref=blog.paperspace.com#part-1-how-bert-is-applied-to-question-answering>
2. Chatgpt

第二題

➤ 第一組(sample code 的 prompt v.s. 空的 prompt)

請從最後一篇的文章中找出最後一個問題的答案	
<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2007年1月16日 <input type="text"/> 完全正確</p> <p>3 1, 200公里, 200公里 <input type="text"/> 接近正確</p> <p>4 2, 廣州, 首都。</p> <p>5 3, 2015, 1990</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 1911年10月10日武昌起義</p> <p>10 8, 香港, 香港</p> <p>11 9, 1921, 2015</p> <p>12 10, 7月, 2010年7月</p> <p>13 11, 2008年, 廣州在1990年以前已經獲得了從「創建國家健康城市」獲得的獎</p> <p>14 12, 從化市, 廣州的人口密度是廣州市的百倍分之一以下。</p> <p>15 13, 80, 非洲黑人</p> <p>16 14, 非洲黑人, 非洲黑人</p> <p>17 15, 普通話, 廣州地區的歷史與文化是廣東人與外國人共同的歷史與文化</p>	<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2014年6月</p> <p>3 1, 200公里, 200</p> <p>4 2, 廣州, 首都</p> <p>5 3, 2015, 1990年</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 1911年10月10日</p> <p>10 8, 香港, 廣州解放紀念碑周圍</p> <p>11 9, 1921, 2012年12月1日</p> <p>12 10, 7月, 2010</p> <p>13 11, 2008年, 廣州獲得「創建國家健康城市」獎項是從1990年開始的。</p> <p>14 12, 從化市, 廣州的人口密度是100倍分之一以下。</p> <p>15 13, 80, 嚴格</p> <p>16 14, 非洲黑人, 廣州</p> <p>17 15, 普通話, 廣州地區的廣東話是廣東話。廣州地區的廣東話是</p>
<p>根據產生的結果，我發現空的 prompt，也就是不下任何指令，雖然 model 還是能回答出一些正確答案(完全回答正確三題)，但整體來說其答對率成斷崖式下降，且回答的答案幾乎跟問題沒有關聯。</p>	

➤ 第二組(sample 的 prompt v.s. sample 翻譯成英文的 prompt)

請從最後一篇的文章中找出最後一個問題的答案	Please find the answer to the last question from the last article
<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2007年1月16日 <input type="text"/> 完全正確</p> <p>3 1, 200公里, 200公里 <input type="text"/> 接近正確</p> <p>4 2, 廣州, 首都。</p> <p>5 3, 2015, 1990</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 1911年10月10日武昌起義</p> <p>10 8, 香港, 香港</p> <p>11 9, 1921, 2015</p> <p>12 10, 7月, 2010年7月</p> <p>13 11, 2008年, 廣州在1990年以前已經獲得了從「創建國家健康城市」獲得的獎</p> <p>14 12, 從化市, 廣州的人口密度是廣州市的百倍分之一以下。</p> <p>15 13, 80, 非洲黑人</p> <p>16 14, 非洲黑人, 非洲黑人</p> <p>17 15, 普通話, 廣州地區的歷史與文化是廣東人與外國人共同的歷史與文化</p>	<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2007年1月16日</p> <p>3 1, 200公里, 200公里</p> <p>4 2, 廣州, 廣州</p> <p>5 3, 2015, 1990年</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 廣州文章: 廣州是中國大陸最早的一個大城市, 當時廣州的</p> <p>10 8, 香港, 廣州</p> <p>11 9, 1921, 廣州地區的歷史與歷史文章: 廣州地區的歷史與歷史文章</p> <p>12 10, 7月, 2010年7月</p> <p>13 11, 2008年, 香港</p> <p>14 12, 從化市, 老城區</p> <p>15 13, 80, 80</p> <p>16 14, 非洲黑人, 非洲黑人</p> <p>17 15, 普通話, 廣州地區的大型公共場所, 普通話是主要交流語言。在某些情況下</p>
<p>根據產生的結果，我發現就算把 prompt 改稱英文，model 回答的準確率其實差不多，用英文的 prompt 模型也能看懂指令，且兩者當中還有一題的答案回答的是一樣，因此我認為將 prompt 用英語的方式呈現並不會對模型有太大的區別。</p>	

➤ 第三組(sample 的 prompt v.s. 改寫的 prompt)

請從最後一篇的文章中找出最後一個問題的答案	不要透過文章，請直接回答問題
<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2007年1月16日 完全正確</p> <p>3 1, 200公里, 200公里 接近正確</p> <p>4 2, 廣州, 首都。</p> <p>5 3, 2008, 1990</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 1911年10月10日武昌起義</p> <p>10 8, 香港, 香港</p> <p>11 9, 1921, 2015</p> <p>12 10, 7月, 2010年7月</p> <p>13 11, 2008年, 廣州在1990年以前已經獲得了從「創建國家健康城市」獲得的獎</p> <p>14 12, 從化市, 廣州的人口密度是廣州市的自倍分之一以下。</p> <p>15 13, 80, 非洲黑人</p> <p>16 14, 非洲黑人, 非洲黑人</p> <p>17 15, 普通話, 廣州地區的歷史與文化是廣東人與外國人共同的歷史與文化</p>	<p>1 ID, Ground-Truth, Prediction</p> <p>2 0, 2007年1月16日, 2007年1月16日</p> <p>3 1, 200公里, 200公里</p> <p>4 2, 廣州, 廣州</p> <p>5 3, 2008, 1990年</p> <p>6 4, 文化大革命, 文化大革命</p> <p>7 5, 廣州, 廣州</p> <p>8 6, 廣州, 廣州</p> <p>9 7, 武昌起義, 1911年10月10日武昌起義</p> <p>10 8, 香港, 香港和台灣</p> <p>11 9, 1921, 廣州文章: 廣州是廣州的首府。廣州是廣州的省</p> <p>12 10, 7月, 2010年7月</p> <p>13 11, 2008年, 廣州獲得從1990年開始參與的「創衛」是從1992年。文章</p> <p>14 12, 從化市, 廣州的人口密度是100倍分之一以下。</p> <p>15 13, 80, 80%</p> <p>16 14, 非洲黑人, 非洲黑人</p> <p>17 15, 普通話, 廣州的普通話與廣東話是主要交流語言。</p>
<p>根據產生的結果，我發現就算讓 prompt 不根據文章回答內容，model 依舊會根據文章並回答答案，且兩個 prompt 所表現出的結果並無太大的差異，準確度差不多。我認為會有這樣的情況是因為 model 在學習的過程中，就已經學到要透過文章然後回答問題，因此在 prompt 中，有沒有加根據文章回答之類的句子並不會對結果有太大的影響。</p>	