**研究動機**

自然語言處理是結合人工智慧和語言學領域的重要方向，注重於自然語言與電腦間的通訊互動。主要包含自然語言理解及自然語言生成，其中自然語言理解從自然語言的表達及語句中識別出該句的目的及含義。然而，機器理解過程常因段落句子組合不同而識別出不同的含義，包含一詞多義造成的理解錯誤。現行的模型訓練主要以自然語言理解 (NLU, Natural Language Understanding) 開發預訓練模型 (Pre-trained Model)，以輔助人機之間的溝通理解及後續自然語言生成 (NLG, Natural Language Generation)，因此語義特徵之提取能力尤為重要。

自然語言生成有能夠完成問答、閱讀理解、段落總結的能力。

以遞迴神經網路 (RNN, Recurrent Neural Networks) 做特徵提取，提取過程中，詞依時序讀入被分配不同的權重。然而，隨著詞與詞之距離拉遠及網路深度增加，先前被輸入詞的權重將被稀釋，造成越前面被輸入之資訊量會越來越少。以Transformer作為主要特徵提取模型架構，用以避免先後輸入造成的權重稀釋問題。使用多頭注意力機制針對當前預測詞將同時用到前面和後面的詞進行計算。Transformer透過平行計算提升訓練及計算效率，RNN則因時序複雜度太大導致計算效率低下。  
 Transformer與RNN在單詞處理的主要差別在於時序輸入，Transformer的無序輸入隨著段落句子及單詞組合不同而識別出不同含義，此時加入位置編碼 (Position Encoding) ，將位置編碼與詞嵌入向量內積作為輸入的嵌入向量，使詞向量包含位置訊息。輸入過程將嵌入向量透過多層的多頭注意力機制、前饋層 (Feed-Forward Neural Network)、與層標準化 (Layer Normalization)進行向量維度轉換 。輸出過程則以殘差連接連接多層的多頭注意力與前饋層。

Transformer之特徵提取架構已相當成熟，其中注意力機制不斷經過改善以提升運算效能，以SHA-RNN (Single-Headed Attention RNN)實作注意力達到單頭注意力的技術受到矚目。

Transformer架構中用到多頭注意力 (Multi-Head Attention)，由於無法確認有效運算頭數，造成Transformer存在記憶體瓶頸與短期文本的短期依賴問題。

相較而言，SHA-RNN在每個隱藏層中只保留單個頭的注意力，過程中有效減少多餘運算量。因此，為了提升運算效能，本研究藉SHA-RNN中的最佳化技術Boom Layer與Transformer中的編碼器進行修改重組Transformer架構執行自然語言文本分類任務。

--------------------------------------------------------------------------------------------------------

**研究問題**

Transformer的運算過程因為賦予單一句子中出現重複詞擁有相同權重，造成無法給予相對鄰近詞較大的權重，進而在自然語言理解出現問題。本研究藉由Transformer及 SHA-RNN重組編碼器和解碼器架構，解決 Transformer弱於捕獲短期文本的依賴問題，並以單頭注意力提升運算效率。

本研究使用重組後之編碼器進行更準確的自然語言文本分類任務。首先，對輸入語句以基於Transformer的模型(BERT、RoBERTa、XLNet、DistilBERT)進行預訓練，包含詞塊化 (Tokenization)、文本清理 (Text Cleaning) 及模型訓練 (Model Pre-training)。訓練過程以比BERT擁有更多訓練語料的模型RoBERTa透過SHA-RNN降低運算量，實作注意力最佳化以解決Transformer的短期文本依賴問題，與傳統的下映射層相比能減少整個矩陣運算量。

因此，本研究將對Transformer的編碼器、SHA-RNN注意力機制與多頭注意力機制之不同、Transformer與SHA-RNN的融合效果進行分析和探討。