**基於Transformer及遞歸神經網路的自然語言情感分析方法**

**摘要**

本研究的目的是以具單頭注意力的遞歸神經網路SHA-RNN (Single-headed Attention RNN)改良Transformer解碼器中多頭注意力機制 (Multi-headed Attention)並與原本Transformer編碼器結合的過程。本研究來源以BERT (Bidirectional Encoder Representation from Transformers) 、SHA-RNN及Transformer中的自注意力機制為主軸進行研究。研究中所使用的方法是透過改造後的遞歸神經網路實現單頭注意力，取代Transformer解碼器中的多頭注意力機制，以改良Transformer在長文本序列中弱於捕獲文本中的短期依賴問題。因此，新的模型架構能透過SHA-RNN及原本的Transformer編碼器，重組編碼器和解碼器架構。新的模型與原本的Transformer和SHA-RNN相比能同時滿足長文本序列輸入所需的短期依賴及具備Transformer原有的長期依賴特點。本研究所獲得的研究成果可以輔助應用於情感分析、社交網路分析及疾病傳播預測。

**A Sentiment Analysis Method based on Transformer and Recurrent Neural Network**

**Abstract**  
The purpose of this research is to implement Single-headed Attention, and replace Transformer decoder with SHA-RNN. The source of this research is based on BERT (Bidirectional Encoder Representation from Transformers), SHA-RNN (Single Headed-Attention with RNN), and the Transformer Self-Attention mechanism as the main axis. The method proposed in this research is to realize the single-headed attention mechanism through the modified RNN, then replaces the self-attention mechanism in the Transformer decoder to improve the problem that Transformer self-attention mechanism is weak in capturing short-term dependence in long sequence. As a result, the architecture of the encoder and decoder can be reorganized through the RNN with a single-head attention mechanism and the original Transformer encoder. Compared with the original Transformer and SHA-RNN, the new model can both meets the short-term dependence required for long sequence input and possesses the original long-term dependence characteristics of Transformer. The result of this research is developed for improving specific tasks well by applying to sentiment analysis, social media analysis, and disease spreading prediction.

* RNN

遞歸神經網路在處理序列的過程分有編碼器和解碼器，透過將輸入語句編碼成固定維度的中間向量，再做為解碼器的輸入。但由於中間向量維度受限，且RNN只具備短期記憶，無法適用長文本序列所需的長期依賴性。

The Recurrent Neural Network model has an encoder and a decoder architecture. In the process of processing sequence, the input sequence would be encoded into a fix-sized vector, which is used as the input of the decoder. However, due to the fixed-size of the intermediate vector, RNN only possesses with short-term memory, it cannot apply for the long-term dependencies required by long text sequence.

* LSTM

LSTM與RNN相比，具備較長的短期記憶，但記憶容量受限於維度和記憶體大小。LSTM和RNN在編碼器的編碼過程，都依序讀取整個序列再轉換成中間向量，因此都有前項單詞權重被稀釋的問題和無法滿足更長序列所需的長期依賴。

Compared with RNN, LSTM has a long short-term memory, but the capacity of memory is limited by the dimensions and memory size. Both LSTM and RNN read the entire sequence in order in the encoding process, and convert it into the intermediate vector. Therefore, both LSTM and RNN have the problem that the weights of previous words are diluted, and the long-term dependencies cannot meet the longer dependencies for longer sequence.

* Transformer

Transformer和LSTM相比，能夠對特定單詞分配權重，而不會因為順序讀取造成的權重稀釋，因此，Transformer能滿足長序列文本所需的長期依賴。Transformer包含編碼器和解碼器部分，但在解碼的過程中，會因為有太多個頭造成的額外計算負擔和額外的運算時間。

Compared with LSTM, Transformer is able to assign weights to specific words without the dilution of weights caused by sequential encoding. Therefore, Transformer can meet the long-term dependencies required by long sequence text. Transformer contains an encoder and a decoder, but in the decoding part, there will be an extra calculation and time cost due to too many heads in the training process.

* BERT

BERT是Transformer的編碼器，透過遮罩和預測句子間的關係做預訓練，能夠在編碼部分做得比以往的模型還要好。此外，也繼承Transformer的特性，能透過平行處理減少多餘的運算時間。

BERT is encoder of Transformer, pre-trained by Masked Language Model, and Next Sentence Prediction, can do better than proposed model. In addition, BERT also inherits the characteristics of Transformer, which can reduce extra time cost and redundant calculation through parallel processing.

* SHA-RNN

Transformer雖然能滿足長文本序列所需的長期依賴，但在短期依賴方面表現得沒有比LSTM好。此外，Transformer在做訓練時，有多個頭一起進行訓練，但無法得知哪些頭真正對訓練有幫助，因此也沒辦法透過減少頭的數量來降低計算負擔。SHA-RNN以改良後的LSTM達到單頭注意力，在模型上的表現能夠節省額外的計算成本。而且在短期記憶方面，也繼承LSTM的優點，能在短期依賴方面表現得很好。

因為在Transformer編碼器的部分，BERT已經超越以往模型的表現，因此本研究透過將SHA-RNN取代Transformer解碼器使新模型能同時在長期和短期依賴方面表現得更好。

Although Transformer can meet the long-term dependencies required for long text sequences, it does not perform better than LSTM in short-term dependencies. In addition, there are multiple heads calculated during the training process, but we cannot know how many heads is actually used in the process. Thus, there is impossible to reduce the extra computation by reducing the number of heads during the training process. Single-headed Attention with RNN (SHA-RNN) implements single-headed attention with the improved LSTM, which saves the extra computation costs. In terms of short-term memory, SHA-RNN also inherits the advantages of LSTM, which can perform well on short-term dependencies.

Since BERT has performed well in the encoder part of Transformer, we proposed a new model which replaces the decoder of Transformer with SHA-RNN, and enable the new model to perform better in both long-term and short-term dependencies.