文獻回顧

卷積神經網路（Convolution Neural Network, CNN）在近代的深度學習模型當中，多被應用於圖像辨識，例如LeCun於1998年所發表的[1]，奠定了卷積神經網路的重要基礎；Alex所提出的多層卷積神經網路架構AlexNet[2]，奪得了2012年的ImageNet冠軍；以及考慮了殘差，進而被創造出來的ResNet[3]，更是終結了ImageNet這場行之有年的圖像辨識比賽。雖說如此，由於卷積神經網路的根本概念是將資料點p的資訊與其周圍k個資訊納入考量，此一特性若運用於時間序列資料上，亦會考慮時間數列在時間點t與前後n個時間點彼此之間的時間相關性，故使用卷積神經網路對語音資料進行處理的研究亦不在少數。在卷積神經網路尚未十分成熟時，便有[4]所提出的類卷積神經網路，運用於語音辨識上的實際例子，近幾年也有利用卷積神經網路對語音進行辨識的研究[5]。

由於考慮到時間前後的相關性，亦有多位研究者投入使用循環神經網路（Recurrent Neural Network, RNN），作為時間序列資料預測模型的核心技術。雖說循環神經網路能夠考慮時間序列資料的時間性，但缺點在於其記憶時間十分短暫，也因此長短期記憶網路（Long Short-Term Memory, LSTM）[6]，作為循環神經網路的改良版，更加廣泛的被應用於實務上。多層的神經網路所堆疊的架構[7][16]，往往能帶來更為優良的結果，這在ImageNet比賽上已獲得了證明[2][8][3]，因此近代的長短期記憶網路，常會使用多層的長短期記憶網路架構[9]，而其往往也會有比較好的結果。此外，雙向的長短期記憶網路（Bidirectional LSTM, BiLSTM）[10]，亦為長短期記憶網路的重要架構，其於Seq2Seq上的應用[11]，一再地表明了同時使用向前與向後的時間序列資料，能使得模型更加準確。

若談論到自然語言處理（Natural Language Processing, NLP），不得不提到Encoder-Decoder模型[12]。Encoder-Decoder模型在提出後，迅速獲得高度關注，由於其可以將一不固定長度的時間數列資料投射至一固定長度的時間數列資料上[9]，且能透過梯度下降法（Gradient Descent）同時進行模型訓練，因此相對於以往的循環神經網路，Encoder-Decoder能使時間數列資料有更多發展性。隨後，Attention模型[13][14]隨著Encoder-Decoder模型的問世，因運而生。由Bahdanau與Luong先後提出的Attention模型架構，使自然語言處理獲得了重大的躍進。該架構先是使用了Encoder-Decoder模型作為基礎，再將Encoder的各個輸出的資訊集中彙整，並作為輸入的一部份使用於Decoder之中，解決了長短期記憶網路在過長的時間序列上，時間數列資料訊息消失問題。日後，Attention被應用於[15]，並且很好的對齊了語音資料與各個英文字母，結果十分顯著，再次檢證了Attention模型在Encoder-Decoder模型上的重要性。

[1] Gradient-Based Learning Applied to Document Recognition

[2] ImageNet Classification with Deep Convolutional

[3] Deep Residual Learning for Image Recognition

[4] Phoneme Recognition Using Time-Delay

[5] Convolutional Neural Networks for Speech Recognition

[6] LONG SHORT-TERM MEMORY

[7] Training and Analysing Deep Recurrent Neural Networks

[16] How to Construct Deep Recurrent Neural Networks

[8] Going Deeper with Convolutions

[9] Sequence to Sequence Learning with Neural Networks

[10] Bidirectional recurrent neural networks

[11] Bidirectional LSTM-CRF Models for Sequence Tagging

[12] Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

[13] NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

[14] Effective Approaches to Attention-based Neural Machine Translation

[15] Attention Is All You Need