

使用長短期記憶類神經網路建構中文語音辨識器之研究

A study on Mandarin speech recognition using Long Short- Term

Memory neural network

賴建宏 Chien-hung Lai

國立交通大學電信工程研究所

Institute of Communications Engineering

National Chiao Tung University

lsr950082@speech.cm.nctu.edu.tw

王逸如 Yih-Ru Wang

國立交通大學電機工程學系

Department of Electronic Engineering

National Chiao Tung University

yrwang@cc.nctu.edu.tw

摘要

近年來類神經網路(Neural network)被廣泛運用於語音辨識領域中，本論文使用遞迴式類神經網路(Recurrent Neural Network)訓練聲學模型，並且建立中文大辭彙語音辨識系統。由於遞迴式類神經網路為循環式連接(Cyclic connections)，應用於時間序列訊號的模型化(Modeling)，較於傳統全連接(Full connection)的深層類神經網路而言更有益處。

然而一般單純遞迴式類神經網路在訓練上隨著時間的遞迴在反向傳播(Backpropagation)更新權重時有著梯度消失(Gradient vanishing)以及梯度爆炸(Gradient exploding)的問題，導致訓練被迫中止，以及無法有效的捕捉到長期的記憶關聯，因此長短期記憶(Long Short-Term Memory, LSTM)為被提出用來解決此問題之模型，本研究基於此模型架構結合了卷積神經網路(Convolutional Neural Network)及深層類神經網路(Deep Neural Network)建構出 CLDNN 模型。

訓練語料部分，本研究使用了 TCC300(24 小時)、AIShell(162 小時)、NER(111 小時)，並加入語言模型建立大辭彙語音辨識系統，為了檢測系統強健度(Robustness)，使用三種不同環境之測試語料，分別為 TCC300(2.4 小時，朗讀語速)、NER-clean(1.9 小

時，快語速，無雜訊)、NER-other(9 小時，快語速，有雜訊)。

關鍵詞：遞迴式類神經網路、長短期記憶、梯度消失(爆炸)、聲學模型、中文、大辭彙
語音辨識、卷積類神經網路、深層類神經網路

Abstract

In recent years, neural networks have been widely used in the field of speech recognition. This paper uses the Recurrent Neural Network to train acoustic models and establish a Mandarin speech recognition system. Since the recursive neural networks are cyclic connections, the modeling of temporal signals is more beneficial than the full connected deep neural networks.

However, the recursive neural networks have the problem of gradient vanishing and gradient exploding in the backpropagation, which leads to the training being suspended. And the inability to effectively capture long-term memory associations, so Long Short-Term Memory (LSTM) is a model proposed to solve this problem. This study is based on this model architecture and combines convolutional neural networks and deep neural networks to construct the CLDNN models.

Keywords: RNNs, LSTMs, gradient vanishing (exploding), acoustic model, Mandarin, LVCSR, CNNs, DNNs