基於字元階層之語音合成用文脈訊息擷取

Character-Level Linguistic Features Extraction

for Text-to-Speech System

陳冠宏　Kuan-Hung Chen

Department of Electronic Engineering, National Taipei University of Technology
s970428@gmail.com

廖書漢　Shu-Han Liao

Department of Electronic Engineering, National Taipei University of Technology
sam8105111@gmail.com

廖元甫　Yuan-Fu Liao

Department of Electronic Engineering, National Taipei University of Technology
yfliao@mail.ntut.edu.tw

王逸如　Yih-Ru Wng

College of Electricl and Computer Engineering, National Chiao-Tung University
yrwang@mail.nctu.edu.tw

摘要

優良的語言文脈訊息是語音合成的關鍵部分,傳統的文脈訊息都是依賴於自然語言處理 (Natural Language Processing,NLP),使用 parser 分析文字。但是 parser 設計困難無法專門為語音合成設計;所以我們想直接以字元為處理單元建立一個 end-to-end 的語音合成系統,在這想法下我們改用字元層級(character-level)的 word2vec 與遞迴類神經網路,直接將輸入字元序列轉換成隱藏特徵向量當做語言合成的文脈訊息。最後我們利用一中英夾雜語音合成系統測試此想法,語音合成的實驗的結果表明,我們提出的方式的確比傳統使用 parse 的方式有更好的性能。

關鍵詞:語音合成、語言特徵、文脈訊息

參考文獻

[1] Stanford-Parser：http://nlp.stanford.edu/software/lex-parser.shtml , 2016, July

[2] Deep Learning in NLP：http://licstar.net/archives/328 , 2016, July

[3] Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber, LSTM: A Search Space Odyssey, eprint arXiv:1503.04069, 03/2015

[4] X Zheng, H Chen, T Xu, Deep Learning for Chinese Word Segmentation and POS Tagging, EMNLP, 2013 - aclweb.org

[5] Chuang Ding , Lei Xie ; Jie Yan ; Weini Zhang ; Yang Liu, Automatic prosody prediction for Chinese speech synthesis using BLSTM-RNN and embedding features, 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp. 98 – 102, 2015

[6] 黃昭銘, 林葭華, 改善條件隨機域模型於中文斷詞, 2010

[7] Eric Brill, A SIMPLE RULE-BASED PART OF SPEECH TAGGER ,1992

[8] 唐大任, 王逸如, 中文斷詞之研究, 2002

[9] Tomas Mikolov, Ilya Sutskever, Kai Chen,Greg Corrado, Jeffrey Dean, Distributed Representations of Words and Phrases and their Compositionality,2013

[10] Tomas Mikolov ∗,Geoffrey Zweig, Context dependent recurrent neural network language model,2012