The 2016 Conference on Computational Linguistics and Speech Processing
ROCLING 2016, pp. 1-3

評估尺度相關最佳化方法於華語錯誤發音檢測之研究
Evaluation Metric-related Optimization Methods for Mandarin Mispronunciation Detection

許曜麒 Yao-Chi Hsu, 楊明翰 Ming-Han Yang, 洪孝宗 Hsiao-Tsung Hung,
林奕儒 Yi-Ju Lin, 陳柏琳 Berlin Chen
國立台灣師範大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan Normal University
{ychsu, mh_yang, alexhung, lin_yj, berlin}@ntnu.edu.tw

陳冠宇 Kuan-Yu Chen
中央研究院資訊科學研究所
Institute of Information Science
Academia Sinica
kychen@iis.sinica.edu.tw

摘要

全球化時代來臨，為提升個人的競爭力，外語能力已列為基本的技能之一。因此電腦輔助語言學習(computer assisted language learning, CALL)在現今已是相當具有潛力的研究；其目的是透過電腦自動判斷外語學習者的學習狀況並給予有幫助的回饋。語言學習又分為聽(listening)、說(speaking)、讀(reading)和寫(writing)等四類學習面向，而本篇論文將專注於電腦輔助發音訓練(computer assisted pronunciation training, CAPT)，也就從是「說」的技術進行討論。電腦輔助發音訓練最主要目的就是要讓第二外語(second-language, L2)學習者有更多的機會練習發音；過去第二外語學習者要進行發音練習都需要配合語言教師的授課時間，若將電腦輔助發音訓練普及到現有的智慧型行動裝置，將會有更多的第二外語學習者因此受惠。電腦輔助發音訓練的首要任務正是錯誤發音檢測，其目的是請學習者讀誦口說教材，針對學習者念誦的錄音，標記學習者的發音是正確發音(correct pronunciation)或錯誤發音(mispronunciation)，標記的目標可以是音素(phone)層次[1]、音節(syllable)層次[2]或詞(word)層次[3]。當系統指出學習者的錯誤發音時，將可以針對該錯誤發音進行偏誤回饋，該階段被稱為錯誤發音診斷[4][5][6][7][8]。近年來，在語音辨識系統中的聲學模型已由深層類神經網路(deep neural network, DNN)取代傳統的高斯混合模型(Gaussian mixture model, GMM)，並在語音辨識任務上取得巨大的進步[9]。在錯

誤發音檢測的相關研究中也因為深層類神經網路聲學模型的使用而在效能上有顯著的提升[10][11][12]。基於上述研究的啟發，我們延續過去學者以最大化錯誤發音檢測任務的效能[13][14]為目標函數對模型進行調整的想法，並實作於深層類神經網路聲學模型的架構上探討對於錯誤發音檢測任務的影響。本論文的貢獻大致可分為三點：1) 比較不同的發音分數做為錯誤發音檢測的評估依據，並探討對於錯誤發音檢測效能的影響；2) 並以發音檢測任務之效能做為更新模型參數的目標函數，實驗顯示將會大幅提升錯誤發音檢測的效能；3) 使用F_1度量作為目標函數時，若將二類的F_1度量線性組合並調整權重，可有效處理資料類別不平衡的問題。本論文的實驗建立在臺灣師範大學邁向頂尖大學計畫所錄製的華語學習者口語語料庫，內容為外國人學習華語所錄製的單字、單詞與短句。從實驗結果可以發現以最大化F_1度量為目標對模型的參數進行調整，在錯誤發音檢測任務上的效果可以得到顯著的提升。

關鍵詞：電腦輔助發音訓練、錯誤發音檢測、自動語音辨識、鑑別式訓練與深層類神經網路

致謝

本論文之研究承蒙教育部－國立臺灣師範大學邁向頂尖大學計畫(104-2911-I-003-301)與行政院科技部研究計畫(MOST 104-2221-E-003-018-MY3 和 MOST 105-2221-E-003-018-MY3)之經費支持，謹此致謝。

參考文獻

[1] S. M. Witt and S. J. Young, “Phone-level pronunciation scoring and assessment for interactive language learning,” *Speech Communication*, vol. 30, no. 2–3, pp. 95–108, 2000.

[2] F. Zhang, C. Huang, F. K. Soong, M. Chu, and R. H. Wang, “Automatic mispronunciation detection for Mandarin,” in *Proc. ICASSP*, 2008.

[3] L. Y. Chen and J. S. R. Jang, “Automatic pronunciation scoring with score combination by learning to rank and class-normalized DP-based quantization,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11 pp. 787–797, 2015.

[4] A. M. Harrison, W. Y. Lau, H. Meng and L. Wang, “Improving mispronunciation detection and diagnosis of learners’ speech with context-sensitive phonological rules based on language transfer,” in *Proc. Interspeech*, 2008.

[5] A. M. Harrison, W. K. Lo, X. J. Qian and H. Meng, “Implementation of an extended recognition network for mispronunciation detection and diagnosis in computer-assisted pronunciation training,” in *Proc. SLaTE*, 2009.

[6] W. K. Lo, S. Zhang and H. Meng, “Automatic derivation of phonological rules for mispronunciation detection in a computer-assisted pronunciation training system,” in *Proc. Interspeech*, 2010.

[7] Y. B. Wang and L. S. Lee, “Improved approaches of modeling and detecting error patterns with empirical analysis for computer-aided pronunciation training,” in *Proc. ICASSP*, 2012.

[8] Y. B. Wang and L. S. Lee, “Supervised detection and unsupervised discovery of pronunciation error patterns for computer-assisted language learning,” *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 23, no. 3, pp. 564–579, 2015.

[9] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition,” *IEEE Transactions on Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[10] W. Hu, Y. Qian, F. K. Soong and Y. Wang, “Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers,” *Speech Communication*, vol. 67, pp. 154–166, 2015.

[11] X. Qian, H. Meng and F. K. Soong, “The use of DBN-HMMs for mispronunciation detection and diagnosis in L2 English to support computeraided pronunciation training,” in *Proc. Interspeech*, 2012.

[12] W. Hu, Y. Qian and F. K. Soong, “A DNN-based acoustic modeling of tonal language and its application to Mandarin pronunciation training,” in *Proc. ICASSP*, 2014.

[13] H. Huang, H. Xu, X. Wang and W. Silamu, “Maximum F1-score discriminative training criterion for automatic mispronunciation detection,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5 pp. 787–797, 2015.

[14] Y. C. Hsu, M. H. Yang, H. T. Hung and B. Chen, “Mispronunciation detection leveraging maximum performance criterion training of acoustic models and decision functions,” in *Proc. Interspeech*, 2016.