

使用語音評分技術輔助台語語料的驗證

Using Speech Assessment Technique for the Validation of Taiwanese Speech Corpus

李毓哲 Yu-Jhe Li¹, 王崇喆 Chung-Che Wang¹, 陳亮宇 Liang-Yu Chen²,
張智星 Jyh-Shing Roger Jang³, 呂仁園 Ren-Yuan Lyu⁴

¹國立清華大學資訊工程學系, ²國立清華大學資訊與應用研究所,

³國立臺灣大學資訊工程學系, ⁴長庚大學資訊工程學系

[yuje.li@mirlab.org](mailto:yujhe.li@mirlab.org), geniusturtle@mirlab.org, davidson.chen@mirlab.org,
jang@mirlab.org, renyuan.lyu@gmail.com

摘要

本論文的主要研究為使用語音辨識及結合語音評分,對未整理的台語語料進行初步的篩選。藉由機器先過濾掉有問題的音檔,如錄音音量過小、太多雜訊、錄音音檔內容有誤等情形,取代傳統人工聽測費時的作法。本論文如圖一所示,可分為三個階段,分別是:「基礎聲學模型訓練」、「語音評分與錯誤原因標記」及「效能評估」。

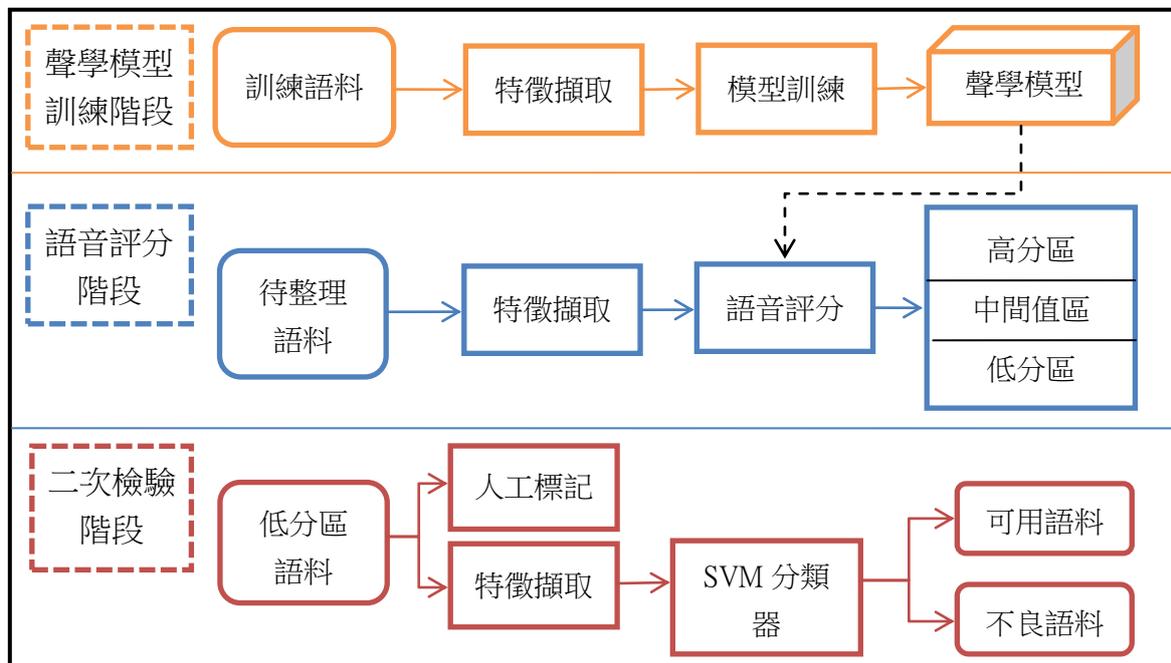
於基礎聲學模型訓練階段,以長庚大學提供的台語語料 ForSD (Formosa Speech Database) [1] 為材料,使用隱藏式馬可夫模型 (Hidden Markov Model, HMM)、梅爾倒頻譜係數 (Mel-frequency Cepstral Coefficients, MFCCs) [2] 和對數能量 (Log energy) 做為語音特徵進行聲學模型的訓練。聲學模型單位分別為:單音素聲學模型 (Monophone acoustic model)、音節內右相關雙連音素聲學模型 (Biphone acoustic model) 及音節內左右相關三連音素聲學模型 (Triphone acoustic model),其針對測試語料進行自由音節解碼辨識網路 (Free syllable decoding) 的音節辨識率 (Syllable accuracy) 最佳結果分別為:27.20%、43.28%、45.93%。其中左右相關三連音素聲學模型的辨識率最佳,因此我們選擇此模型進行第二階段的實驗。

於語音評分與錯誤原因標記階段,將於基礎聲學模型訓練階段已訓練好的左右相關三連音素聲學模型,對待整理的語料進行語音評分 [3, 4]。語音評分能藉由聲學模型對錄音進行評分,在本論文中以評分後的分數來評量音檔的與文本間的相似程度。但依據前人研究 [5],在某些狀況下語音評分的分數並不合理,因此在本論文中,為了降低評分時不合理情形出現的機率,加入了三種分數調整的扣分機制,分別是:音節之音框個數差距過大、音節中連續音素之音框數目過小、以及文本與辨識結果之音節數目不一。而此評分結果將依照門檻值分為三部分,分別為低分區、中間值區及高分區。且針對低分區部分語料進行人工標記,標記其錯誤原因,再對其擷取特徵,使用支持向量機 (Support Vector Machine, SVM) 訓練出分類器,最後以該分類器對低分區語料進行二次檢驗,將低分區語料分為可用語料及不良語料。

於效能評估階段,將原先訓練語料分別加入「未整理語料」、「中間值區及高分區語料」、「高分區語料」進行聲學模型的訓練,比較篩選語料前、後效能,其音節辨識率結果分

別為：40.22%、41.21%、44.35%。

由結果看來，經過篩選後語料所訓練出的聲學模型與未經篩選語料所產生的聲學模型，其辨識率的差別最高可達 4.13%，證實本論文所提的方法，藉由語音評分確實能有效的自動篩選掉有問題的語句。



圖一、語料整理系統流程圖

關鍵詞：台語語料整理、隱藏式馬可夫模型、語音評分、語音辨識、支持向量機

Keywords: Taiwanese Corpus Validation, Hidden Markov model, Speech Assessment, Support Vector Machine.

參考文獻

- [1] Ren-yuan Lyu, Min-siong Liang, Yuang-chin Chiang, Toward Construction A Multilingual Speech Corpus for Taiwanese (Min-nan), Hakka, and Mandarin, International Journal of Computational Linguistics and Chinese Language Processing, 2004.
- [2] Steven B. Davis and Paul Mermelstein, Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences, IEEE International Conference on Acoustics, 1980.
- [3] 李俊毅，語音評分，清華大學碩士論文，民國 91 年。
- [4] 陳宏瑞，使用多重聲學模型以改良台語語音評分，清華大學碩士論文，民國 100 年。
- [5] 黃武顯，基於 32 位元整數運算處理器之華語語音評分的改良與研究，清華大學碩士論文，民國 96 年。