

A Pilot Study on Automatic Scoring of GEPT High-Intermediate Speaking Tests

高照明

台灣大學外國語文學系

摘要

本研究根據 400 份全民英檢中高級口語測驗的音檔與成績以統計及機器學習演算法實做自動英語口語能力評分。我們採用計算語言學工具將評分的程序自動化，透過軟體自動將音檔轉成文字並產生 51 個與英語口語能力相關的特徵。本研究嘗試回答下列 5 個問題。(1). 哪一個特徵與得分的相關係數最高？(2). 利用這 51 個特徵作為自變項的迴歸模型其解釋力如何？(3). 將分數分成過與不過兩種，邏輯迴歸能解釋多少百分比的資料？(4). 以監督機器學習演算法來預測實際得分正確率有多高？(5). 使用監督式機器學習演算法預測通過與否的正確率有多高？實驗結果發現與分數關聯性最高的都是 ngram 相關的特徵。其中與分數的關聯係數最高的是受試者在測驗時所使用且在 COCA 語料庫出現的 5 連詞 (5grams) 的數目，達到 0.3568。這個特徵加上 type/token ratio 這個特徵的邏輯迴歸能正確預測 67.8% 是否能通過測驗。使用機器學習演算法，有 98.76% 的預測分數與原始分數的差距在一分之內。這顯示本研究所採用的方法可以用於諸如能力分班的低風險的測驗或用於線上練習。