

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

Abstract

This study aims to investigate how well test-takers' oral proficiency can be modeled by statistics and machine learning algorithms based on 400 sample audio files taken from the GEPT high-intermediate speaking tests. Procedures have been automated to transcribe the audio files, generate computational indexes, and score the test-takers' responses. There are five research questions addressed in this study: (1). Which computational index correlates most strongly with score? (2). How well can step-wise regression account for the relationship between computational indexes and scores? (3). How well can the computational indexes and logistic regression account for the data on a pass/fail basis? (4). How well can supervised machine learning algorithms perform in predicting the scores? (5). How well can supervised machine learning algorithms perform on a pass/fail basis? In total, 51 features have been identified. Of the 51 features, the features regarding the number of ngrams attested in the COCA reference corpus have been found to correlate with scores more strongly than any other types of features. In addition, in general the longer the ngrams attested in COCA, the higher the correlation coefficients. The highest is 5gram, whose correlation coefficient with scores is 0.3568. The logistic regression with the number of 5grams attested in the COCA reference corpus and the type/token ratio can account for 67.8% of the data. While the proposed machine learning algorithms and the features cannot predict the actual scores with high accuracy, 98.76% of the predicted scores are within one point from the original scores. This suggests that the proposed approach might be used in low-stake tests such as placement tests or online writing exercises.