Cambridge English Centenary
Symposium on Speaking Assessment

Automated assessment: Moving from written text to transcribed speech

Ted Briscoe, University of Cambridge

The task of automated assessment (AA) of free text focuses in the EFL context on analysing and assessing the quality and variety of writing competence. Automated assessment systems exploit textual features in order to measure overall quality and assign a score to a text. The earliest systems used superficial features, such as word and sentence length, as proxies for examiners' judgements. More recent systems have used more sophisticated automated text processing techniques to measure grammaticality, textual coherence, pre-specified errors, and so forth.

Deployment of AA systems gives a number of advantages, such as reduced workload in marking texts, especially when applied to large-scale assessments. Additionally, automated systems guarantee the application of the same marking criteria, thus reducing inconsistency, which may arise when more than one human examiner is employed. Often, implementations include feedback with respect to the writers' writing abilities, thus facilitating self-assessment and self-tutoring.

Most work has treated AA as a supervised text classification task, where training texts are labelled with a grade and unlabelled test texts are fitted to the same grade point scale via a regression step applied to the classifier output and texts are represented in terms of manually pre-specified features. In our work with Cambridge English Language Assessment we have treated AA as a supervised discriminative machine learning problem where the task is to rank scripts on an ordinal scale and the features used to represent the text are selected and weighted automatically as part of the training process. This removes the need for the regression step better modelling the grading task, and also removes the need to manually pre-specify the aspects of the text that are assumed to be criterial for grading and recoverable using current language processing technology.

I will describe and motivate our approach to grading written text and report experimental results which suggest that our AA system's performance is essentially indistinguishable from a human examiner when applied to text similar to that seen during training. I will then go on to describe the challenges of applying similar AA techniques to the grading of speech. Primary amongst these are the issues of transcribing L2 speech produced by speakers of varying abilities and L1 backgrounds.

However, I will focus on the applicability of our language processing and AA technology to transcribed output and here I will attempt to demonstrate that the outlook is very promising.