Off-topic detection systems for spoken language assessment

Vatsal Raina, Mark Gales

ALTA Institute/Department of Engineering, Cambridge University

{vr311,mjfg}@cam.ac.uk

Abstract

There is growing interest in automatic spoken language assessment and teaching systems for English as demand to learn English increases. "Free speaking" tasks, where the candidates are required to speak spontaneously in response to a prompted question, is a standard mode of spoken language assessment. For these tests it is important to identify when the candidate's response is not relevant to the prompt, as this may indicate a lack of understanding of the question or malpractice. This work contrasts previously published topic relevance systems with transformer-based models, as well as assessing the impact of combining these forms system together. As there is usually limited manually annotated data to train these systems, a self-supervised training approach based on permuting questions and responses is adopted for all systems. Experiments were conducted on the Cambridge Assessment English Linguaskill Business and General English tests covering candidate responses from a broad range of abilities in English proficiency with CEFR levels from A1 to C2. A couple of general trends are observed for all systems: when the prompts are seen during training system performance is significantly better than for unseen prompts; and systems perform better on artificial, permuted, test data than "real" manually annotated off-topic data. Across the models, the transformer-based models are more effective at identifying off-topic responses to unseen prompts, and combine well with existing approaches. Additionally, it is shown that only a small number of prompt/response pairs are required per prompt for fine-tuning such models to achieve substantial performance gains.