A PROMPT-AWARE NEURAL NETWORK APPROACH TO CONTENT-BASED SCORING OF NON-NATIVE SPONTANEOUS SPEECH

Yao Qian, Rutuja Ubale, Matthew Mulholland, Keelan Evanini, Xinhao Wang

Educational Testing Service Research, USA

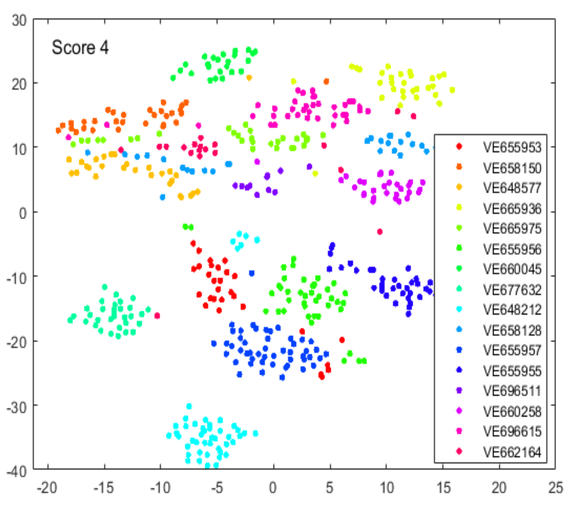
Measuring the Power of Learning.®

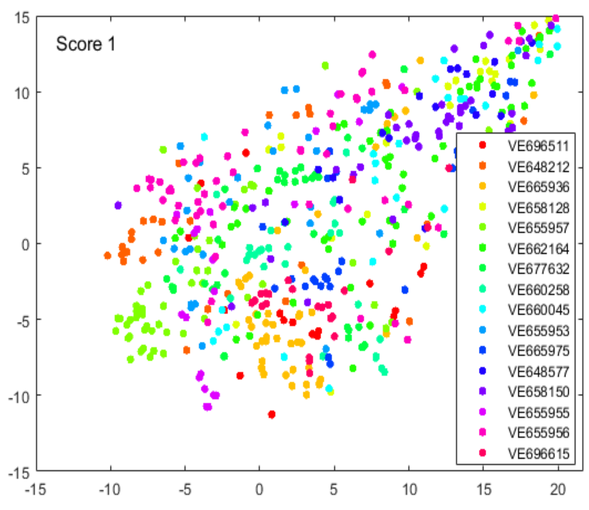
Summary

- We proposed a prompt-aware NN approach to build a generic scoring model without doing any feature engineering.
- It performs as well as the strong baseline of a Support Vector Regressor using content-related features.
- It is more effective in grading responses to unseen prompts.

Response Visualization Using t-SNE

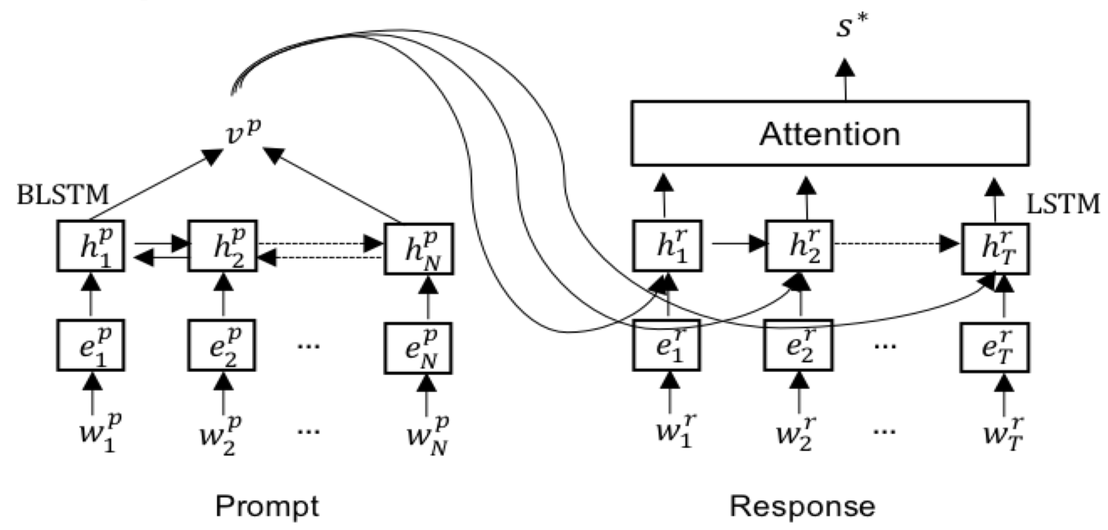
- Each point represents a response and its color distinguishes different prompts





Model Architecture

- Prompt-aware and attention-based LSTM-RNN for scoring spoken responses



Data and Baseline

- Data Partition

Partitions	Speakers	Responses	Prompts
Train	2,511	6,635	16
Test(Seen/Unseen)	714(635/79)	2,103(1,870/233)	18(16/2)

- Baseline: SVR with content features from C-rater
  - Character n-grams for n=2 to 5
  - Word unigrams and bigrams
  - Length of response in characters
  - Syntactic dependencies
  - Prompt bias

Experimental Results

- Performance across the NN structures (w/) or (w/o) filtering model and different NN structures

NN Structures	Correlations
Attention LSTM-RNN (w/o)	0.789
Siamese LSTM-RNN + Attention LSTM-RNN (w)	0.800
Prompt encoder + Attention LSTM-RNN (w/o)	0.806

NN Structures	Correlations
LSTM-RNN	0.739
+WE trainable	0.745
+Attention layer	0.789
+Prompt encoder	0.806

- Performance across different systems

	Baseline	Att_RNN	Att_RNN_P
Human, All	0.801 (0.769)	0.789 (0.758)	0.806 (0.767)
Human,Seen	0.813 (0.761)	0.804 (0.764)	0.815 (0.773)
Human,Unseen	0.770 (0.729)	0.735 (0.694)	0.773 (0.737)
ASR, ALL	0.767 (0.727)	0.782 (0.754)	0.791 (0.769)
ASR, Seen	0.794 (0.771)	0.799 (0.773)	0.798(0.774)
ASR, Unseen	0.731 (0.659)	0.701 (0.632)	0.731(0.675)

- Performance across different responses to seen prompts