Research Week 2021

Fine-tuning BERT for the semantic classifier of the PNT

Alexandra C Salem, Robert Gale, Steven Bedrick
salem@ohsu.edu
Computer Science & Electrical Engineering, OHSU

Keywords

aphasia, speech, machine learning, natural language processing

Abstract

Background

Aphasia is a language disorder resulting from a stroke or brain injury. Individuals with aphasia make paraphasias: speech errors consisting of an unintended word. In the Philadelphia Naming Test (PNT), a participant with aphasia is shown a picture (e.g. a table), and asked to say what it is. If the participant makes a paraphasia, the clinician classifies the participant's response (e.g. "chair") into six categories based on lexicality, phonological similarity to the target (table), and semantic similarity to the target.

Objectives

This work focuses on automatically classifying semantic similarity of the paraphasia. We train a modern language model called BERT alongside a binary classifier to categorize each transcribed response to a PNT item as semantically similar to the target or not. We evaluate the accuracy of this classification compared with clinician scores. We also compare the results of this model to a previously used language model called word2vec.

Methods

Our dataset is a subset of the Moss Aphasia Psycholinguistic Database (MAPPD) consisting of 11,999 clinician-transcribed and categorized responses to PNT items. We also use a supplementary dataset of context sentences for each target word. We run several experiments to train BERT with the classifier (with and without context sentences) and compare our results with previous methods.

Results

The previous classifier using word2vec achieves 0.9389 F1 (0.9474 precision, 0.9306 recall) for the not similar class and 0.8394 F1 (0.8201 precision, 0.8597 recall) for the similar class. Our best-performing BERT model achieves 0.9496 F1 (0.9548 precision, 0.9444 recall) for the not similar class and 0.8657 F1 (0.8532 precision, 0.8786 recall) for the similar class.

This corresponds to BERT making 880 errors instead of 1,062 errors. Finally, we conduct an error analysis between the two models.

Conclusions

Our best-performing BERT model improves automatic classification of the semantic similarity of PNT target-response pairs.