2019 methodology recap

Andrew Yang

July 31, 2020

1 Goal

To devise and implement an automated pipeline for generating word-duration and pitch track variables to predict specialists' suggestions/interventions for Moby examples.

2 Baseline

First we cluster specialists' free-form observations and suggestions into categories. With these observations and suggestions, we use a random forest machine learning model as a baseline for prediction. We spawned a separate random forest to predict each suggestion.

The model performed well (> 80% accuracy) in a small handful of categories, but there was more than enough room for improvement

3 Featurization

In 2019 we proposed generating more features to supplement (and to replace, if possible) the observations. We were motivated by the following considerations:

- A very limited dataset
- Inadequacies of observations in predicting interventions

We dictated that these additional features be automatically generated.

Preprocessing word durations

We mean-centered word durations for all examples (including gold).

Preprocessing pitch tracks

We centered F0 frames by subtracting from each frame the mean F0 across the example. For each token in each example, we then compute the pitch slope ((end F0 - start F0) / nframes).

6 Aligning examples to gold

Aligning examples to source text

Using the standard DIFF procedure we aligned each example to the source text (recstring).

Pause-delimited alignment

Absent a more sophisticated method for identifying meaningful constituencies, for each gold example to we identify **pause-delimited** groups of tokens.

For each of these groups of tokens, we generate **vectors** consisting of word durations and pitch slopes.

7 Similarity score

Pseudocode explains this better. For a given example:

```
for each Gold Reading G:
  for each pause-delimited group of tokens P:
    E := tokens from the diff alignment corresponding to P
    V_E := vectors for the duration and pitch track for E
    V_P := vectors for the duration and pitch track for P
    Scores := compute cosine similarities between vectors in V_E and V_P
    store Scores
```

8 Additional featurization

Some additional features were generated as well, including the method described above after removing all pauses, as well as weighted cosine similarity scores by the **length** of the tokens.

9 Preliminary results

We applied the same random forest model to our combined set of features to predict the same interventions as before. We observed a much higher raw prediction accuracies for most of the intervention categories. We also observed a much higher correlation between the automatically generated features and the intervention categories, as opposed to the specialists' observations we used in the baseline.