The Importance of Sub-Utterance Prosody in Predicting Level of Certainty

Heather Pon-Barry
School of Engineering and Applied Sciences, Harvard University
Cambridge, MA 02138, ponbarry@eecs.harvard.edu

Stuart Shieber
School of Engineering and Applied Sciences, Harvard University
Cambridge, MA 02138, shieber@seas.harvard.edu

Overview
We address the problem of predicting the perceived level of certainty of a spoken utterance. We have a corpus of utterances spoken under varying levels of certainty. In each utterance, a single word or phrase is responsible for the speaker’s level of certainty. We investigate whether using prosodic features of this word or phrase and of its surrounding context improves the prediction accuracy when compared to using features taken only from the utterance as a whole.

We go beyond previous work by looking at the predictive power of prosodic features extracted from salient sub-utterance segments. Previous research on uncertainty has examined the predictive power of utterance- and intonational phrase-level prosodic features (Liscombe et al., 2005). Our results suggest that we can do a better job at predicting an utterance’s perceived level of certainty by using prosodic features extracted from the whole utterance plus ones extracted from salient pieces of the utterance, without increasing the total number of features, than by using only features from the whole utterance.

This work is relevant to spoken language applications in which the system can identify locations likely to cause uncertainty. Examples of such systems include tutorial dialogue systems (Pon-Barry et al., 2006; Forbes-Riley et al., 2008) and second language learning and literacy systems (Alwan et al., 2007).

Prosodic Features

The Combination feature set (shaded in table below) is created by selecting either the whole utterance feature, the context feature, or the target word feature, whichever one is most strongly correlated with perceived level of certainty.

Uncertainty Corpus

- 20 speakers
- 600 utterances

Method of elicitation:
1. Speakers are presented with a sentence containing one or more gaps
2. Options for filling in the gap are displayed
3. Upon hearing a beep the speaker reads the sentence aloud

Vocabulary

Transportation
Q: How can I get from Harvard to the Silver Line?
A: Take the Red Line to ______.
  a. South Station
  b. Downtown Crossing

Vocabulary
Only the ______ workers in the office laughed at all the manager’s bad jokes.
  a. pugnacious
  b. craven
  c. sycophantic
d. spoffish

Take the Red Line to South Station.

Target Word

Context

Five annotators rated the perceived level of certainty on a 5-point scale
Speakers rated their own level of certainty on the same 5-point scale

Linear Regression model accuracies
- Value to predict: perceived level of certainty
- Results shown: 20-fold ‘leave one speaker out’ cross-validation averages

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Num Features</th>
<th>Accuracy (5 classes)</th>
<th>Accuracy (3 classes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Utterance</td>
<td>20</td>
<td>39.00%</td>
<td>54.25%</td>
</tr>
<tr>
<td>(B) Target Word</td>
<td>20</td>
<td>43.14%</td>
<td>68.96%</td>
</tr>
<tr>
<td>(C) Context</td>
<td>20</td>
<td>37.11%</td>
<td>67.50%</td>
</tr>
<tr>
<td>(D) All</td>
<td>60</td>
<td>48.54%</td>
<td>74.58%</td>
</tr>
<tr>
<td>(E) Combination</td>
<td>20</td>
<td>45.42%</td>
<td>74.79%</td>
</tr>
</tbody>
</table>

Are the differences due to noise?

Combination set predictions are more strongly correlated with perceived level of certainty than Utterance set predictions in 16 out of 20 folds

Acknowledgements

This work was supported in part by a National Defense Science and Engineering Graduate Fellowship.

Self vs. Perceived Level of Certainty

Self-reported levels of certainty were consistently lower than perceived levels of certainty