IMPROVING INTONATIONAL PHRASING WITH SYNTACTIC INFORMATION

Philipp Koehn, Steven Abney, Julia Hirschberg, Michael Collins

AT&T Research, 180 Park Ave., Florham Park, NJ 07932
`koehn@isi.edu, (abney,julia,mcollins)@research.att.com`

ABSTRACT

The prediction of intonational phrase boundaries from raw text is an important step for a text-to-speech system: Locating where to place short pauses enables more natural sounding speech, that can be more easily understood. We improved upon earlier work [Hirschberg and Prieto, 1996] by adding syntactic information gained from a high-accuracy parser [Collins, 1999]. We report significant improvement using various experimental setups. We also show that our improved method comes close to interannotator agreement.

Introduction

In spoken language, intonational phrases are separated by breaks in form of pauses in the speech. Consider the following sentence:

`A broken tax pledge by a Democratic incumbent | is an issue in West Virginia`

In spoken language, a native speaker is likely to pause for a short interval between the words `incumbent` and `is` (the break is indicated in the sentence above by an horizontal bar). She is unlikely to pause between, for instance, `tax` and `pledge`, and the sentence read this way would be less comprehensible. Thus, in a good text-to-speech system, a proper placement of these intonational phrase breaks is important.

In our approach we consider that a pause may occur between any pair of adjacent words. We try to train a decision structure to predict if a word is followed by a break. A manually labeled corpus of 68,011 words is used for experimentation.

We build on top of earlier work [Hirschberg and Prieto, 1996], which used a set of features consisting of: a 4-word POS window, 2-word window for pitch accents, the total number of words and syllables in the utterance, the distance of the word from beginning and end of the sentence in words, syllables, and stressed syllables, distance from the last punctuation in words, the type of any punctuation that directly follows the word, whether the word is at the end, within, or at the beginning of an NP, and if within an NP, its size and the distance of the word from the start of the NP. We call this feature set the *old* feature set, which we will extend.

This feature set was trained on an implementation of CART [Breiman et al., 1984], a decision tree learner. We report results on the old and our new feature sets with CART, before we move on to other machine learning algorithms.

Syntactic Features

We acquire syntactic information by parsing the sentence with a high accuracy syntactic parser [Collins, 1999]. It produces a parse tree[1], such as:

```
TOP S NP-A NPB DT A
               JJ broken
               NN tax
               NN pledge
           PP IN by
              NP-A NPB DT a
                       JJ Democratic
                       JJ incumbent
      VP VBZ is
         NP-A NPB DT an
                  NN issue
              PP IN in
                 NP-A NPB NNP West
                          NNP Virginia
```

The break in our example sentence occurs after the 8-word `NP-A` phrase, which is followed by 6-word `VP` phrase. Our intuition is that intonational phrase breaks occur between large syntactic phrases. It is also more likely to have an intonational phrase break after a major phrase (`NP`, `VP`, `PP`, `ADJP`, `ADVP`).

This leads us to define three syntactic features:

- the size of the biggest phrase that ends with this word
- a binary flag indicating if the phrase is a major phrase
- the size of the next phrase on that level in the parse tree

If the next phrase is just one word (like the conjunct `and` or the auxiliary verb `is`), we collapse it with the following phrase with regard to these features.

Since intonational phrase breaks are quite common before sub-clauses, we added a fourth feature:

- a binary flag indicating if the following phrase has the label `SBAR`

[1] The parser also produces head word information, which is not used in this work.

Results

We evaluated these features on a corpus containing 60456 words for training and 7555 words for testing. Experimental results are summarized in the following table:

feature set	precision	recall	F
4 new	81.7%	62.4%	70.8%
old	90.1%	77.0%	83.0%
old + 3 new	88.9%	80.1%	84.3%
old + 4 new	89.3%	80.8%	84.8%

Using the all the new features, we could improve the f-measurement[2] by 1.8%. Also note the effectiveness of the fourth syntactic feature.

When examining the decision tree, we find the added syntactic features very high in the tree, meaning that they are the first features the classifier considers while making its decision. Clearly, they provide effective discriminating information.

Other machine learning algorithms confirm this finding. We used Boostexter [Schapire and Singer, 1999], a boosting algorithm on simple features, Ripper [Cohen, 1995], a rule learner, C5.0, a boosted decision tree classifier[3], and an alternating decision tree method [Freund and Mason, 1999] with strikingly similar results.

Adding other features, such as the actual words around the break or syntactic labels did not improve results. Even cheating, by allowing the classifier to know the true position of the last break preceding the position under consideration, did not help significantly.

Learning Curve

Providing different amounts of training data, we can observe the following learning curve:

training set size	precision	recall	F
10,000 ex.	86.6%	52.7%	65.5%
20,000 ex.	86.9%	68.9%	76.8%
30,000 ex.	87.5%	77.1%	82.0%
40,000 ex.	88.2%	78.3%	83.0%
50,000 ex.	89.4%	79.6%	84.2%
60,456 ex.	90.1%	80.0%	84.8%

This indicates that more data should improve results.

Optional Breaks and Interannotator Agreement

Placing intonational phrase boundaries is not clearly defined. There are positions in the sentence where it lies mostly in the subjective judgement of the speaker to pause. However, at most positions either there must or must not be a break. In a new labeling scheme we considered this and introduced an *optional* tag for positions, where no clear decision could be made.

[2]The f-measurement (F) is computed from precision (P) and recall (R) by F = 2RP/(R+P)

[3]C5.0 was written by Ross Quinlan, it is an extension of C4.5. The program is available commercially at `http://www.rulequest.com`

For training and testing we ignored optional breaks. Hence, the classifier learned to make clear decisions and the performance numbers are higher. Also, since we relabeled only a small portion of the corpus this way (less than 20,000 words), we expect a boost from more training data.

Having the data labeled twice also allowed us to measure interannotator agreement, since the different labeling was done by different people. In the following table, the annotations of human A are taken to define truth, and human B is considered as a classifier.

classifier	precision	recall	F
human B	95.3%	90.9%	93.0%
old features	88.2%	84.1%	86.1%
old + 4 syn	92.1%	87.4%	90.8%

We make two observations. First, the improvement afforded by syntactic features in the detection of intonational phrase boundaries is reconfirmed (90.8% vs. 86.1%). Second, the trained classifier performs almost as well as the human.

Conclusions

We have demonstrated that adding syntactic features improves the quality of intonational phrasing within our framework. Also, a larger training set size should deliver better results. Adding more detailed syntactic or lexical information, however, did not have a major impact. Neither did the use of different machine learning methods.

Comparing our results with interannotator agreement, we seem to be approaching the limit of what such a system could possibly achieve.

REFERENCES

[Breiman et al., 1984] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, P. J. (1984). *Classification and Regression Trees.* Wadsworth Internation Group.

[Cohen, 1995] Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning.*

[Cohen and Singer, 1999] Cohen, W. W. and Singer, Y. (1999). Simple, fast, and effective rule learner. In *AAAI.*

[Collins, 1999] Collins, M. (1999). *Head-Driven Statistical Model for Natural Language Parsing.* PhD thesis, University of Pennsylvania.

[Freund and Mason, 1999] Freund, Y. and Mason, L. (1999). The alternating decision tree learning algorithm. In *Proceedings of the 16th International Conference on Machine Learning.*

[Hirschberg and Prieto, 1996] Hirschberg, J. and Prieto, P. (1996). Training intonational phrasing automatically for english and spanish text-to-speech. *Speech Communication*, 18:281–290.

[Schapire and Singer, 1999] Schapire, R. R. and Singer, Y. (1999). A boosting-based system for text categorization. *Machine Learning (submitted).*