Predicting and Detecting Prosodic Prominence

Dan Jurafsky

Stanford University

I describe a number of experiments with my colleagues on the predicting and detection of pitch accents with the goal of improving tools like automatic speech recognition and synthesis.

In our first study (Brenier et al. 2006, Nenkova et al. 2007, Yuan et al. 2005) we attempt to predict which words in a text string should bear pitch accent, using sentences from the Switchboard corpus. The linguistic literature points out that complex features such as information status, contrast, and animacy help predict pitch accent placement. By using the Switchboard corpus, annotated for such features, we explore whether these rich features improve prominence prediction over traditional shallow features such as frequency and part-of-speech. We show that by far the best predictive feature is a simple lexical one called 'accent ratio' which measure how likely the word was to be accented in some labeled corpus.

In our second experiment (Sridhar et al 2008) we extend our previous study to include acoustic features, to use features of the context (neighboring words), and to use more sophisticated machine learning classifiers. We show the best published performance on the task of pitch accent detection, using a CRF classifier and very local contexts of a single word on either side. The accent ratio of a word is again the best feature, although adding acoustic information provides additional improvements in accent detection accuracy.

In our third experiment (Strom et al. 2007) we attempt to use these predictors to improve the performance of a unit-selection text-to-speech synthesizer, examining three levels of accent: unaccented, pitch accented, and emphatic pitch accented. We first introduce a number of methods for adding emphatic accented data into the unit database used by the synthesizer. We then synthesize sentences using default accent prediction (based only on content-function word differences), using our accent ratio classifier, and using emphatic words marked in the sentences by the author via capitalization. We show that accent prediction and emphatic accent both improve the quality of synthesis, as measured by human preference in a listening test.

- Jason Brenier, Ani Nenkova, Anubha Kothari, Laura Whitton, David Beaver, Dan Jurafsky. 2006. The (Non)Utility of Linguistic Features for Predicting Prominence in Spontaneous Speech. IEEE/ACL 2006 Workshop on Spoken Language Technology, Aruba.

- Ani Nenkova, Jason Brenier, Anubha Kothari, Sasha Calhoun, Laura Whitton, David Beaver, and Dan Jurafsky. 2007. To Memorize or to Predict: Prominence Labeling in Conversational Speech. NAACL-HLT 2007.

- Vivek Kumar Rangarajan Sridhar, Ani Nenkova, Shrikanth Narayanan and Dan Jurafsky. Detecting prominence in conversational speech: pitch accent, givenness and focus. Speech Prosody, Campinas, Brazil, 2008.

- Volker Strom, Ani Nenkova, Robert Clark, Yolanda Vazquez-Alvarez, Jason Brenier, Simon King, and Dan Jurafsky. 2007. Modelling Prominence and Emphasis Improves Unit-Selection Synthesis. Interspeech 2007.

- Yuan, Jiahong, Jason M. Brenier, and Dan Jurafsky. 2005. Pitch Accent Prediction: Effects of Genre and Speaker. In Proceedings of EUROSPEECH-05.