ERROR ANALYSIS AND DISFLUENCY MODELING IN THE SWITCHBOARD DOMAIN

Roni Rosenfeld (CMU) (Project Leader)
Rajeev Agarwal (TI), Bill Byrne (JHU)
Rukmini Iyer (BBN), Mark Liberman (UPenn)
Liz Shriberg (SRI), Jack Unverferth (DoD)
Dimitra Vergyri (JHU), Enrique Vidal (UP Valencia)

1995 JHU Summer Workshop
Project Team Report

The 'Language Modeling for Spontaneous Speech' (LMSS) project team at the 1995 Johns Hopkins Workshop ("LM'95") devoted its resources to addressing three major issues: analysis of the baseline Switchboard system, modeling of conversational speech phenomena, and preliminary investigation of "data bleaching" — a new domain adaptation paradigm.

When analyzing the baseline system, we discovered that a word missing from the lattice tends to cause up to 3 word errors, and that such errors are hard to eradicate even with extremely powerful language models. We quantified the loss incurred when working from N-best lists as opposed to lattices, and consequently decided to use 1000-best lists. We learned that doubling the amount of language training data is likely to reduce the word error rate by no more than 2%. We experimented with rank histograms as an alternative measure of modeling progress. We found that a mismatch between segmentation information in the training and test data is a significant problem, and that it can be overcome to some extent by hypothesizing boundaries; linguistic boundries were found to be more informative than acoustic ones. Segment based error analysis revealed surprisingly that disfluent segments are no more error prone than non-disfluent ones, and long segments are no more erroneous than short ones. Finally, word based error analysis yielded a list of features strongly correlated with word error.

In modeling conversational speech phenomena, we concentrated on modeling speech disfluencies and back-channel cues. We tried three different approaches to disfluency modeling, neither of which reduced the error rate. However, at least one technique appeared quite successful at automatically annotating disfluencies and at extracting conversational speech phenomena from the data stream without degrading the quality of the model (the latter is useful for data bleaching -see below). In back-channel modeling, we statistically characterized the dynamics of a conversation, devised simple filters to automatically detect back channel cues, and used them to partition the training data accordingly.

Finally, we pursued a new paradigm in domain adaptation, "data bleaching", in which idiosyncracies of data from remote domains are removed so as to make the data statistically closer to the target domain. Data from the target domain can be similarly "bleached", except that "inverse bleaching" must eventually be applied to the resulted model. We derived a list of conversational speech phenomena that could benefit from bleaching, and demonstrated the approach on two of them: substitution of Out-Of-Vocabulary words from remote domains, and excision of disfluencies, hedge words and flavor words from conversational speech.