Listen Meeting – October 30th 2006

Gabriel Murray
Progress Report
Overview

- **Research area**: speech summarization in meeting domain
- **Research interest**: using speech-specific characteristics for summarization
  - Rather than just importing techniques from text summarization
- **Looking at three stages**
  - Pre-extraction: e.g., term weighting
  - Extraction: e.g., prosodic correlates of dialogue acts
  - Post-extraction: compressing dialogue acts
A common term-weighting method is \textit{tf.idf}:
\[ \text{tf}\times\text{math.log}(D/d) \]
- \( D \) is total documents, \( d \) is documents containing the term.

So, good at finding technical and rare words, but what about words like “budget”, “design” and “cost”?

We could look at difference of usage among between speakers.
Speaker Weighting

- The hypothesis is that speakers do not use keywords with the same frequency
  - Especially when roles are well-defined
  - We could see which words are used a lot by one or two people and not used little or not at all by the others
- For each word, for each speaker we compute a surprisal based on the usage among the other three speakers
- Then average those four scores
Speaker Weighting

• Of course, sometimes a speaker will just have a word usage not shared by the others
  – But combining this score with *tf.idf* might yield improvement
• First results are positive: it does give significant improvement
• We evaluate by using those scores to then do extractive summarization
Term Weighting Results

- Weighted Precision
- tf.idf
  - Manual: 0.52
  - Asr: 0.51
- Combo
  - Manual: 0.54
  - Asr: 0.53
- Combo is sig. better on both manual and asr
- Neither approach is sig. worse on asr
Extraction

- Have described several systems in the past
  - Unsupervised system using SVD on matrix of prosodic, lexical and structural features
  - Feature-based approach using GMMs
  - LSA sentence representation
- Newest system: exploit speaker status information
  - Who's leading the meeting?
Extraction

• Step One
  − Determine meeting leader – total perc. of time speaking, number of dialogue acts, speaking consistently throughout

• Step Two
  − Find speaker's most important utterances, based on duration and term weighting
  − Create preliminary summary (350 words)

• Step Three
  − Find most similar utterances by other 3 speakers
  − Using LSA representation, as before
Extraction

• 750 word final summary
• Evaluation:
  – For development, weighted precision
  – Ultimately, extrinsic evaluation: task–based evaluation
• Will have preliminary evaluation soon
Compression

- As described before
  - Preserve the pitch contour
- Allows us to create more informative summaries
- Results encouraging so far
- Will apply to AMI now
  - Compare manual and ASR