RESCORING CONFUSION NETWORKS FOR KEYWORD SEARCH
Victor Soto, Erica Cooper, Lidia Mangu, Andrew Rosenberg and Julia Hirschberg
Columbia University, IBM and Queens College/CUNY

Main Findings
Introduction 2-stage cascaded scheme to rescore Confusion Networks (CNs) for Keyword Search in Low-Resource Languages.
1. Rescore CNs to improve error rate of 1-best hypotheses: Using rank-SVM classifier, obtain WER gains between 0.54% and 2.84%.
2. Generate keyword hits from rescored CNs and use logistic regression to detect true hits and false alarms. Gains between 0.45% and 0.9% in MTWV compared to hits generated from unrescored CNs.

Keyword Search Task
- IARPA Babel program: rapid creation of speech technology for a diverse set of languages with only a small amount of training data.
- Spoken keyword search task: identify all exact matches for some set of query terms, provided as text, in a given corpus of speech.
- Metric: Term-Weighted Value
  \[ \text{ATWV}(\theta) = 1 - [P_{\text{min}}(\theta) + \beta \cdot P_{\text{FA}}(\theta)] \] (1)
  \[ \text{MTWV} = \max \text{ATWV}(\theta) \] (2)

Corpora
- IARPA Babel conversational speech, Full (FLP) and Limited (LLP) language packs, with 80 and 10 hours of training speech respectively. Dev and Eval partitions are 10 and 5 hours long.
- Cantonese, Tagalog, Turkish, Pashto, Vietnamese
- Word Confusion Networks are aligned to transcripts to obtain labels for each arc.

Feature Extraction and Feature Selection
- Lexical:
  - Percentile of word frequency
  - Silence and epsilon flag
  - Number of syllables of token, and syllable index of its primary and secondary stress.
- Phonetic:
  - Count of phones
  - Binary features indicating whether word begins/ends in an unvoiced consonant or glottal stop.
- Syntactic Proxies:
  - model M class labels.

Structural:
- Posterior score
- Arc rank, arc rank - bin size ratio, confusion bin size...
- Bin number at the segment and conversation level, distance to previous and next silence and to beginning and end of segment and conversation.

Feature Extraction and Feature Selection
- Lexical:
  - Percentile of word frequency
  - Silence and epsilon flag
  - Number of syllables of token, and syllable index of its primary and secondary stress.
- Phonetic:
  - Count of phones
  - Binary features indicating whether word begins/ends in an unvoiced consonant or glottal stop.
- Syntactic Proxies:
  - model M class labels.

Structural:
- Posterior score
- Arc rank, arc rank - bin size ratio, confusion bin size...
- Bin number at the segment and conversation level, distance to previous and next silence and to beginning and end of segment and conversation.

Rescoring Confusion Networks
- Table: Most prominent features for CN reranking ( LLP ) according to QPFS.

Rescoring Posting Lists
- Posting List: list of hypothetical hits including keyword id, start time, duration and score.

Future Work
- OOV word handling
- Cross-language experiments

SpeechLab@Columbia: http://www.cs.columbia.edu/speech/lab.cgi
SpeechLab@QC CUNY: http://speech.cs.qc.cuny.edu/