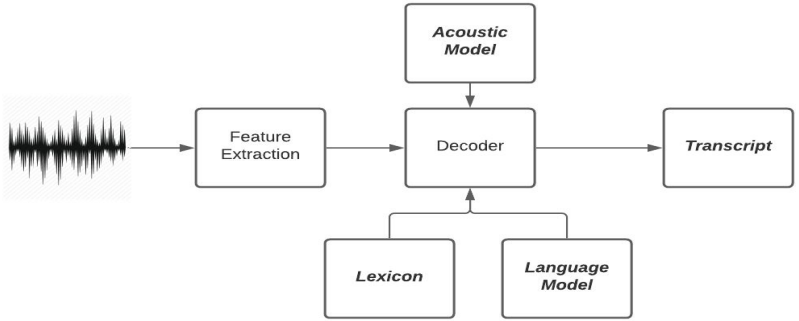
ASR with Punctuation, Capitalization & Numeral

Xudong Liang (Brandon)
UNI: xl2891

Data & Approach



- **Data:**
  - TedLium Release 3 - 452 Hours of Audio, 2351 Aligned Audio TED Talks in NIST (Sphere) Format
  - Singapore National Speech Corpus - 1,000 Hours of Audio, 1,031 Speakers, 826,469 utterances, short speech
- **Approach & Architecture:**
  - Pre-Train on TedLium data (transcript with lower-cased word tokens only) - Kaldi TedLium Recipe
  - Fine-Tune on NSC data (transcript with **Punctuation**, **Capitalization** & **Numerical**) - Mimic Kaldi TedLium Recipe
    - **Lexicon**: use TedLium lexicon as default
      - **Capitalization**: Double/Duplicate original & capitalize 1st letters
      - **Numeral**: Rule-Based (111 = 1 + hundred + and + 11) & Spelled out (111 = 1 + 1 + 1)
      - **Punctuation**: 2 versions
        - 1. Attach to the end of each word token (multiplies Lexicon size by # of unique puncts)
        - 2. As Standalone Tokens (size increase becomes negligible) - NLTK WordPunctTokenizer
    - **Acoustic Model**: Statistical GMM-HMM Model (Mono + Triphone)-> TDNN
    - **Language Model**:
      - Nassar 2020: Transformer LM < RNN LM for small datasets
      - 4-Gram with **3 Lexical Components** -> RNN LM

Numeral Lexicon Mapping - Rule-Based

- Using CMU Pronouncing Dictionary for Special Words & Basic Numbers

Numeral	Type/Rule	ARPAbet Pronunciation
a	Special Word	AH
and	Special Word	AH N D
hundred	Special Word	HH AH N D R AH D
thousand	Special Word	TH AW Z AH N D
1	Basic Number	W AH N
5	Basic Number	F AY V
15	Basic Number	FIH F T IY N
20	Basic Number	T W EH N T IY
25	“20” + “5”	T W EH N T IY F AY V
70	Basic Number	S EH V AH N T IY
75	“70” + “5”	S EH V AH N T IY F AY V
100	“a”/“1” + “hundred”	AH/W AH N HH AH N D R AH D
125	“100” + “and” + “25”	AH/W AH N HH AH N D R AH D AH N D T W EH N T IY F AY V
175	“100” + “and” + “75”	AH/W AH N HH AH N D R AH D AH N D S EH V AH N T IY F AY V
1,000	“a”/“1” + “thousand”	AH/W AH N TH AW Z AH N D
1,175	“1,000” + “175”	AH/W AH N TH AW Z AH N D AH/W AH N HH AH N D R AH D AH N D S EH V AH N T IY F AY V

Test Results - WER

Model	GMM-HMM + 4-Gram	TDNN + RNNLM
TED-LIUM (best)	*16.1%*	*6.7%*
Version 1 (Punct Attached)	39.7%	-
Version 2 (Punct as Standalone Tokens)	**32.5%**	In Progress