Sec 4:

Evaluation

Automatic Metrics

Utterance segmentation

Mitigating error due to speaker variation

Sec 4.1

Automatic Metrics

Evaluation

- Motivated by evaluation in machine translation
  - Automatic evaluation
    - Cheap
    - Fast
  - Human evaluation
    - Gold standard
    - Subjective
    - Expensive, time-consuming

Automatic metrics

- Reuse Text MT-based metrics
  - *BLEU*
    - Compare reference translation to output

- Multi-task system
  - *Word error rate (WER)* of transcription
    - Single correct output
    - Often calculated ignoring punctuation and case

BLEU

- Compare Hypothesis to reference translation
  - Geometric mean of n-gram precision (1 to 4-grams)
  - Using case- and punctuation information

Reference: BLEU is a MT metric

Hypothesis: BLEU is my metric

BLEU

- Compare Hypothesis to reference translation
  - Geometric mean of n-gram precision (1 to 4-grams)
  - Using case- and punctuation information

Reference: BLEU is a MT metric

Hypothesis: BLEU is my metric

1-gram: 3/4
2-gram: 1/3
3-gram: 0/2
4-gram: 0/1

BLEU = $\sqrt[4]{3/4*1/3*0*0*BP}$

BLEU

- Compare Hypothesis to reference translation
  - Geometric mean of n-gram precision (1 to 4-grams)
  - Using case- and punctuation information
- Aggregated scores over large dataset
- "*Brevity penalty*" to account for recall

Reference: BLEU is a MT metric

Hypothesis: BLEU is my metric

1-gram: 3/4
2-gram: 1/3
3-gram: 0/2
4-gram: 0/1

BLEU = $\sqrt[4]{3/4*1/3*0*0*BP}$

Word error rate (WER)

- Align reference and hypothesis
  - Calculate insertions, deletions and substitutions
  - Divide by reference length

Reference: WER is an ASR metric

Hypothesis: WER is my *** metric

- Often ignoring case and punctuation

Word error rate (WER)

- Align reference and hypothesis
  - Calculate insertions, deletions and substitutions
  - Divide by reference length

Reference: WER is an ASR metric

Hypothesis: WER is my *** metric

Alignment: S D

- Often ignoring case and punctuation

Word error rate (WER)

- Align reference and hypothesis
  - Calculate insertions, deletions and substitutions
  - Divide by reference length

- Often ignoring case and punctuation

Reference: WER is an ASR metric

Hypothesis: WER is my *** metric

Alignment: S D

$$\text{WER} = \frac{S+D+I}{N} = \frac{2}{5}$$