KS53
-23

音声認識仮説の曖昧性を考慮するMulti-task End-to-End音声翻訳

○胡 尤佳，須藤 克仁，Sakriani Sakti，中村 哲（NAIST）

[Link] ANLP2021 Paper

研究概要

▷**問題** Multi-task End-to-End音声翻訳(ST)が音声認識出力の曖昧性を考慮できていない

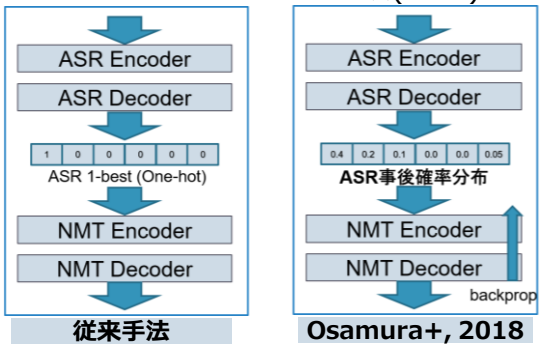
▷**手法** 音声認識(ASR)出力の曖昧性に頑健なモデルの学習方法

▷**結論** ASR事後確率分布を用いた学習による精度の向上

関連研究

▷音声認識出力に頑健なCascade音声翻訳 [Osamura+, 2018]

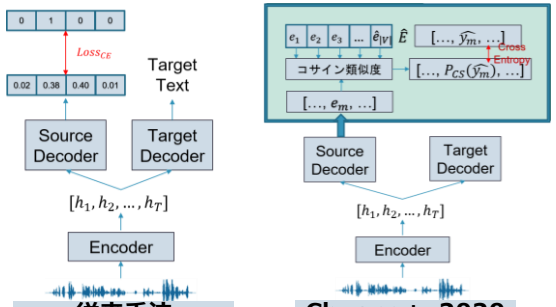
ASR事後確率分布による機械翻訳(NMT)の学習



従来手法　　Osamura+, 2018

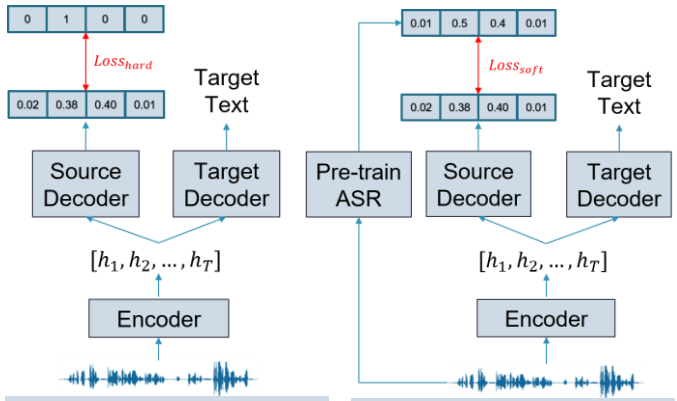
▷意味的類似度に頑健なMulti-task End-to-End音声翻訳 [Chuang+, 2020]

単語分散表現間のコサイン類似度をlossに用いる



従来手法　　Chuang+, 2020

提案手法



従来手法 (hard loss)
参照: One-hot (hard loss)

提案手法 (soft loss)
参照: **ASR事後確率分布**

▷従来手法：One-hotで想定される課題

[参照訳] I catch a ball
[予測パターン①] I **cat** a ball
[予測パターン②] I **lost** a ball

発音の類似度 **高** / **低**

発音の類似度が考慮されない学習になる可能性

▷提案手法：ASR事後確率分布を参照として用いたSTの学習

- ASR事後確率分布が**発音の近い単語同士の類似度の情報を持つ**と仮定
- → **ASR出力の不確実性対して頑健な音声翻訳の実現**を期待

実験

▷**実験条件** コーパス：Fisher CallHome Spanish (Es-En) 実装:ESPnet (Transformer)

▷**ベースライン*：Cross entropy loss**

▷**提案手法*：**Pre-trained ASR：Dev accuracy best

$$\lambda_{ASR} = 0.3, \quad \lambda_{soft} = \{0.0, 0.3, 0.5, 0.7, 1.0\}\ (\lambda_{soft} = 0.0：ベースライン)$$

$$Loss_{ASR} = \lambda_{soft} Loss_{soft} + (1 - \lambda_{soft}) Loss_{hard}$$

$$Loss = \lambda_{ASR} Loss_{ASR} + (1 - \lambda_{ASR}) Loss_{ST}$$

表1:BLEUスコアの比較

	ベースライン	提案手法			
λ_{soft}	0	0.3	0.5	0.7	1.0
Dev	41.04	40.99	41.40	41.20	**41.51**
Dev2	42.14	42.05	42.28	**42.45**	42.22
Test	41.17	41.38	**41.41**	41.18	41.39

(＊記載した実験のLabel smoothingは ST-task：0.0 / ASR-task：0.0)

表2:提案手法($\lambda_{soft} = 0.5$)とベースラインのFisher testの例文比較

Src-ref (Es)	sí pero o sea sigue siendo bastante **intensi**
Tgt-ref (En)	yes but it's still pretty **intensive**
Baseline (En)	yes but that keeps getting pretty unthinkable (→"inconcebible", "impensable" (Es))
Proposed(En)	yes but that keeps being pretty **intense**

▷結果

- soft lossを含めるとBLEUの向上が見られた

▷今後の課題

- ASR-taskの出力結果分析
- Pre-train ASRの性能と翻訳性能の変化の比較
- 発音・読み情報を利用した誤差関数の計算[Salesky+, 2020]