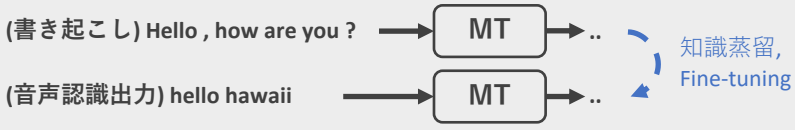
P1-7 人手書き起こしの知識を用いた音声認識誤りに頑健な機械翻訳

○福田 りょう　須藤 克仁　中村 哲
奈良先端科学技術大学院大学 (NAIST)

1. 本研究の概要

本研究では, 機械翻訳 (MT) の学習における
「書き起こし」と「音声認識出力」の効果的な併用について検討



モチベーション

MTの学習に利用できる2種類の入力を考える

「人手による書き起こし」はcleanなテキスト: 高い翻訳精度を獲得
「音声認識出力」はnoisyなテキスト: 認識誤りに対する頑健性を獲得

→ それぞれのいいとこ取りをしたい!

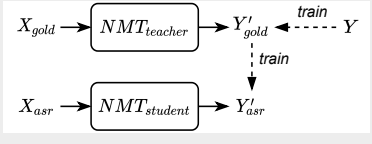
2. 背景

- 音声翻訳システム (SLT): 原言語の音声を目的言語のテキストに変換
  - 音声認識モデル (ASR) と機械翻訳モデル (MT) を持つCascade型が主流
  - Cascade型の問題点: ASRの音声認識誤りが後続のMTへ伝播

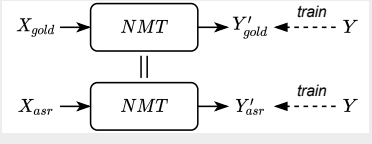
3. 実験設定

- **手法**: 知識蒸留とFine-tuning, またその併用. モデルはTransformer

(a)知識蒸留: 書き起こしで学習した教師モデルの知識を生徒モデルへ蒸留
(b)Fine-tuning: 書き起こし→音声認識出力, の順で単一モデルを学習



(a) 知識蒸留



(b) Fine-tuning

- **データ**: Fisher Spanish-to-English (約14万文)
  - MTの入力として, 書き起こし (Gold transcript) と音声認識出力 (ASR output) (WER 36.5pt) を用いる

4. 結果と考察

学習方式	テストデータ			
	ASR output (ref0/ref1)		Gold transcript (ref0/ref1)	
(1) Gold transcript	17.45	16.98	**26.75**	**26.14**
(2) ASR output	17.49	16.87	17.62	17.15
(3) 知識蒸留	18.48	17.87	16.52	16.24
(4) Fine-tuning	18.31	17.5	24.89	24.52
(5) Fine-tuning + 知識蒸留	**18.76**	**17.96**	25.24	24.86

- ✓ 知識蒸留手法(3)はASR outputに対する翻訳精度がベースライン(1)(2)を有意に上回る. 学習の難しさを緩和する働きを示唆
- ✓ Fine-tuning手法(4)も同様に(1)(2)を上回る.
  また(3)知識蒸留と異なりGold transcriptに対しても精度を維持
- ✓ 両手法の併用(5)が, ASR outputに対しては最も高い翻訳精度.
  Gold transcriptに対しても提案手法のなかで最高
  - Fine-tuningは「パラメータ」を, 知識蒸留は「出力の知識」を書き起こしで学習されたモデルから継承

生成例

Gold transcript	¿Y hace tiempo que ya estás en ésta cosa? ¿De llamar por teléfono?
ASR output	y hace tiempo ya que está en esta cosa de llamar por teléfono
(2)の出力 (ASR)	It's been a long time since you're calling on the phone
(5)の出力 (ASR)	How long have you been in this thing to call on the phone?

- ✓ 音声認識出力に含まれない, 符号や句読点の生成に対しても提案手法は良い働きが期待される.
  上記の例でベースライン(2)は平叙文を訳出した.
  一方で提案手法(5)では訳としてより適切な疑問文を生成できた.

5. 今後の課題

- 音声認識精度による有効性の変化の調査
- 学習手法のより広範な検証
- End-to-End型のSLTとの比較/優位性の強調