S4-P38

方言音声認識における古典的手法の再考

松﨑 孝介 [1*]，谷口 雅弥 [2]　（[1] 東北大学, [2] 理化学研究所）

*matsuzaki.kosuke.r7@dc.tohoku.ac.jp

概要

日本語方言の音声認識において、大規模モデル（Whisper）と古典的手法（CTC）の特徴を比較した

➡ 共通語で一般的でない表現やフィラーにおいて、CTCの有効性を示した

背景

- 音声認識技術は、共通語では高精度を達成しているが**方言に対しては精度が低い**
- 方言の音声認識は、言語・文化の保存において重要
- 大規模音声認識モデル Whisper[1] は、短時間や無音の入力に対して誤出力や反復出力をする傾向がある[2]

目的

- **話したとおりに書き起こす**音声認識手法を構築する

実験方法

- Connectionist Temporal Classification (**CTC**) を用いてスクラッチ学習し、Whisper (large) と比較

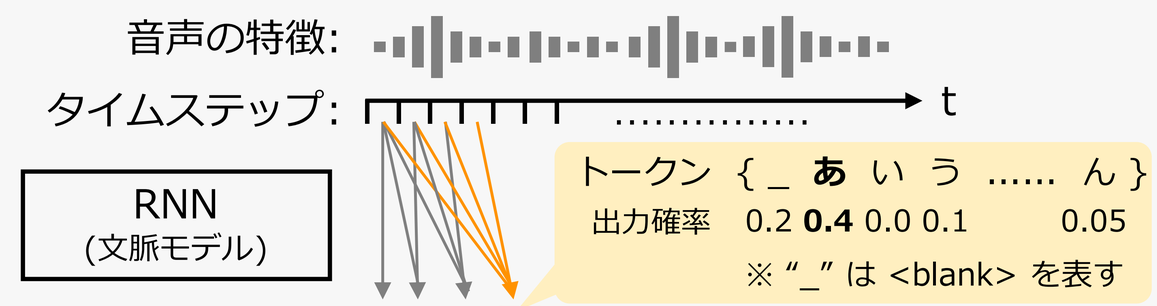

図1　CTCによる音声認識過程

データセット

- 日本語諸方言コーパス（**COJADS**）有償版[3]
  - 音声と書き起こしの組が全国66地点・計105時間
- **Common Voice** Corpus 日本語版[4]
  - 音質が検証されたデータがtrain, dev, testに分割済

表1　各データセットの収録話数・収録時間

	COJADS*1		Common Voice*2	
	発話数	時間 [h]	発話数	時間 [h]
Train	144,147	71.38	15,411	19.97
Dev	17,989	8.88	8,001	9.91
Test	18,028	8.93	8,001	9.74

*1 0.5秒未満の発話を除去し都道府県・発話ごと8:1:1に分割した
*2 GPT-5 により書き起こしテキストをひらがな化した

評価指標

- 文字誤り率（Character Error Rate: **CER**）

$$\mathrm{CER} = \frac{S + I + D}{N} = \frac{\text{Edit Distance}}{N}$$ **（低いほど良い）**

S = 置換文字数, I = 挿入文字数, D = 削除文字数,
N = 正解（参照）テキストの文字数

実験結果

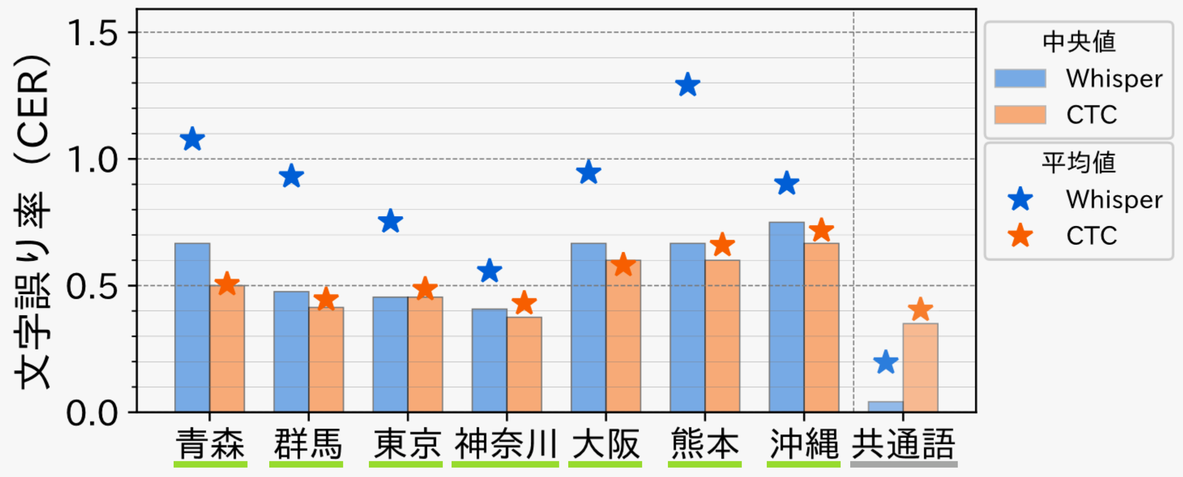

図2　WhisperとCTCの音声認識結果（TestデータのCER）

出力例

特徴 (収録県)	正解・Whisper出力・CTC出力	CER
フィラー (宮城)	ふーん。	-
	ごしちょうありがとうございました	4.00 ☹
	ふーん。	0.00
非共通語的 (兵庫)	ってゆーとっから	-
	というところから	0.75 ☹
	でゆーとっから	0.25
共通語的 (山梨)	みそいれてやいたり	-
	みそをいれてやいたり	0.11
	ねそーゆでてあいたり。	0.67 ☹

同様の例が185件(約1.0%)

考察

COJADS（方言音声）の文字誤り率（CER）

- Whisper中央値 > CTC中央値
  - Whisperは共通語で一般的な表現に寄せる傾向
- Whisper平均値★ ≫ Whisper中央値▬
  - 発話と無関係な出力により高いCERとなる例が存在

両データセットのCTCの性能比較

- COJADSでのCER > Common VoiceでのCER
  - 表2のような特徴の違いが原因と考えられる

表2　各データセットの発話の特徴

	COJADS	Common Voice
内容	全国66地点の方言	共通語
形態	会話（複数話者）	読み上げ（単独話者）
環境	雑音を含む	静音

今後の展望

- データセット間でCTCに性能差が出る要因の分析
- 方言データと共通語データを組み合わせた学習
- 方言語彙辞書の構築と活用

参考文献
[1] Alec Radford *et al*. Robust speech recognition via large-scale weak supervision. 2023.　[2] Mateusz Barański *et al*. Investigation of Whisper ASR Hallucinations Induced by Non-Speech Audio. 2025.　[3] 日本語諸方言コーパス（COJADS, 有償版）Ver.2025.03. https://www2.ninjal.ac.jp/cojads/index.html.　[4] Common Voice Corpus 日本語版 v21.0. https://commonvoice.mozilla.org/ja/datasets.