ロボット設計製作論実習5

非流暢性現象付き発音形音声認識の日常会話音声における評価

藤江研究室 21C1048 小堀 聡太

1. 背景

- 「えー」や「うーん」は人の状態を表す
  - 悩み 緊張の判断
- フィラー検出に時間がかかる
  - システム側の認識問題
    - 検出箇所が限定的
- 従来研究
  - 局所的でなく全体的な音響情報のエンコードを可能にしたブロック処理
  - 日本語日常会話コーパス(CEJC)を利用した言語情報と非言語情報の音声認識

2. 目的

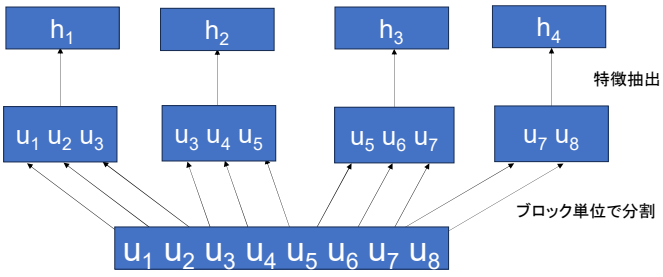
- ストリーミングでのフィラー検出速度向上
  - フィラーを素早く検出
  - 従来のE2Eにストリーミングを追加
- 日常会話での調査
  - 日常場面で自然に生じた会話に対して考察
- 日本語話しコーパス(CSJ)との比較

3. 実験方法

- 実験ツール
  - ESPnet
    - E2Eの音声処理に使用
- コーパス

データベース名	CSJ	CEJC
内容	学会	日常会話

- フィラーに関しての調査方法
  - カナ文字のデータにタグをつけ判断
    - 例:エ+F ーF ナ ン カ イ イ ネ
    - 計算効率化 単純化
- 提案するブロック処理について



ブロックサイズ・・・音声データを一定の長さに分割

ホップサイズ・・・次の音声データの初期位置を決定

- 評価データについて
  - CEJCコーパスから選別
    - 静かな環境の音声
      - 7人の話者によるデータ

4. 実験結果

学習データ	文字誤り率(%)
CSJ	25.2
CEJC	15.3
CSJファインチューニング	15.0

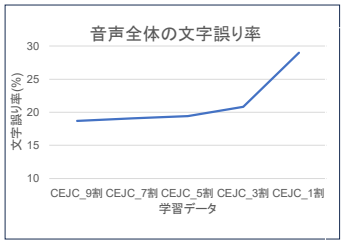
ファインチューニング ・・・ 事前学習モデルを新たなデータを使い再学習する方法

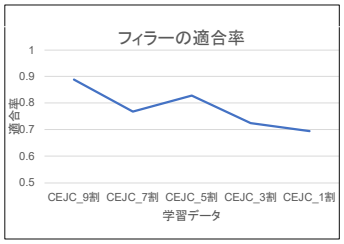
文字誤り率 ・・・システムが認識した結果と正解のテキストとで計算した文字単位の割合

フィラーに関して

学習データ	適合率
CSJ	0.113
CEJC	0.863
CSJファインチューニング	0.864

学習データをCEJCとCSJの混合データに変更





5. 考察

- CSJとCEJCの比較
  - 音声全体について
    - CSJは挿入誤りが多い
      - 間の取り方の違いによる可能性
    - 評価データとの会話内容の違い
      - 文脈の複雑さ
- CSJファインチューニングとCEJCの比較
  - 大部分の文字認識は高いが一部文字で未認識
    - 特定の単語や音声パターンが学習データに不足
- フィラーについて
  - 用途の違い
    - CEJC
      - 話を聞き回答への悩み
    - CSJ
      - 発表内容間の繋ぎ

6. 今後の予定

- 評価データの変更
  - 様々な環境のデータで実験
    - 生活音がある音声
    - 電話の音声

参考文献

[1] Emiru Tsunoo, Yosuke Kashiwagi, Toshiyuki Kumakura, Shinji Watanabe "TRANSFORMER ASR WITH CONTEXUAL BLOCK PROCESSING," ASRU, pp14-18, 2019

[2] 塩根凪人, 若林佑幸, 北岡教英, "言語現象と非言語情報も検出する音声認識システムの提案," 日本音響学会, 2-Q-3, 2023