複数の音声認識器からのシンプルで高精度な認識結果の選択手法

嶋田 和孝・宇津巻 彰（九州工業大学） / NLP若手の会 第3回シンポジウム

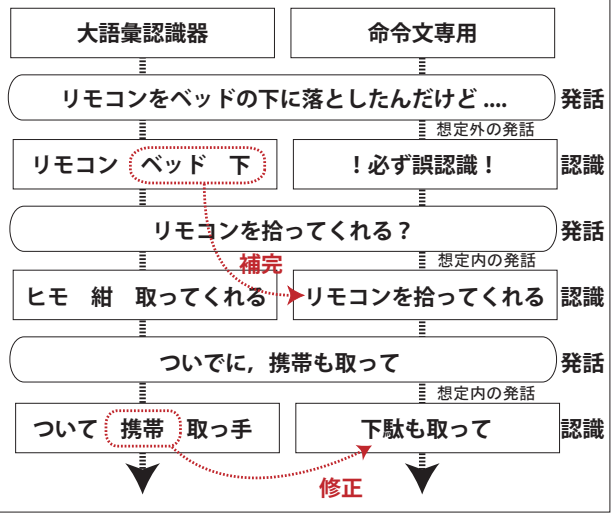
研究の背景・目的

- 介護支援ロボットの開発
  - 音声認識器：認識精度の問題

- 精度向上のアプローチ
  - 複数の認識器を利用
    - タスクごとの認識器の作成と統合
      - カーナビゲーション
        - レストラン検索・ルート検索・観光地検索

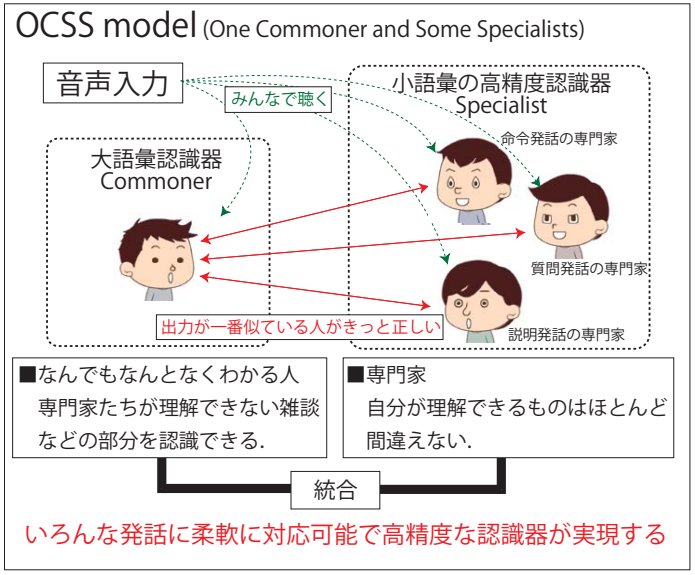
- 複数のタスク依存認識器による複合認識器
  - タスク依存認識器が認識できないものは扱えない
    - 語彙サイズ：小→精度：高・範囲：小
    - 語彙サイズ：大→精度：低・範囲：大

- タスク外の発話
  - 重要な文脈情報になる可能性あり
  - タスク外の発話も柔軟に扱える枠組みの必要性



タスク依存な発話については頑健で，柔軟性の高い手法を実現する

提案手法



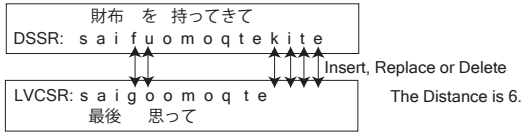
- 提案手法のイメージ -

- 基本的な考え方：誰が最も大語彙認識器と似ているか？
  - 大語彙認識器と似ている＝その認識器の出力が正しい
    - 信頼度の推定などでも用いられる考え方
  - 誰も似ていない＝タスク外発話
- 類似性の判断：2 つの編集距離に基づき判定
  - 発話レベル（ED_{utter}）＆単語レベル（ED_{word}）
  - 最終的な距離：音素数で正規化したもの
  - 分類ルール：順に適用

1) $ED_{utter} < Thresh_{utter}$
   →距離が最小のタスク依存認識器を採用
2) $ED_{word} < Thresh_{word}$
   →距離が最小のタスク依存認識器を採用

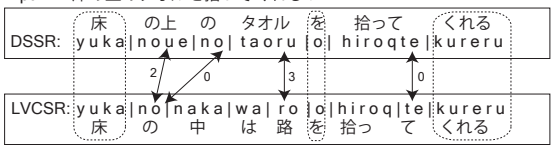
else
→大語彙認識器の結果を採用

* 発話レベル
Input: 財布を持ってきて



ED_{utter} = Distance / # of phonemes in two results
= 6 / 26 = 0.23

* 単語レベル
Input: 床の上のタオルを拾ってくれる？



Best Combinations: noue vs. no, no vs. no, taoru vs. wa|ro, hiroqte vs. hiroq|te
Distance/Phoneme: 2/6=0.33, 0/4=0, 3/9=0.33, 0/14=0
ED_{word} = Sum / # of words in the DSSR result = (0.33+0.33)/7 = 0.09

実験

- 実験内容：認識器を適切に選択できるか？
- 予備実験
  - 2 つの認識器：大語彙認識器（Julius）& 命令発話認識器（Julian）
  - 50 発話（命令）& 50 発話（非命令）× 4 人（男女 2 名ずつ）
  - 交差検定
  - 高い分類精度
  - 揺れない閾値と精度
    - $Thresh_{utter}$: 0.24-0.26, $Thresh_{word}$: 0.08-0.13
    - 閾値固定でも F 値の低下は 0.01 程度

Type	Precision	Recall	F
命令	0.963	0.985	0.974
非命令	0.985	0.963	0.974

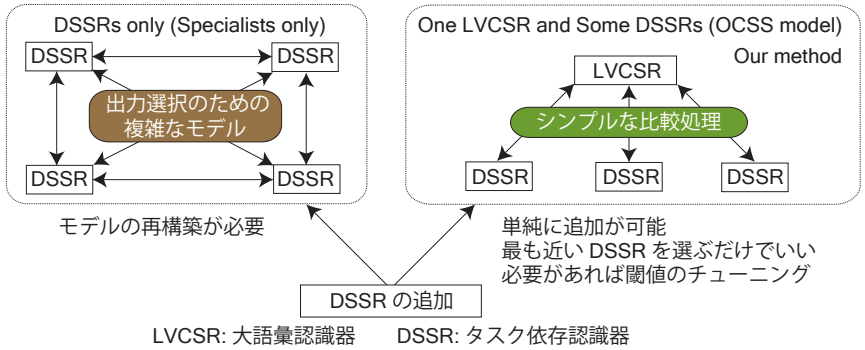
- 本実験
  - 4 つのタスク依存認識器
    - 患者からの命令発話：机の上の携帯を取ってくれる？
    - 看護師からの命令発話：食事を 301 号室に持って行って
    - ロボット制御発話：1m 右に移動
    - 質問発話：TV のリモコンはどこにある？
  - 実験データ
    - 各カテゴリに 10 発話（合計 50 発話）& 被験者：6 名・閾値：予備実験の値

Type	Precision	Recall	F
大語彙認識器	0.838	0.950	0.891
患者からの命令発話	0.983	0.966	0.975
看護師からの命令発話	1.000	0.983	0.992
ロボット制御発話	1.000	0.933	0.966
質問発話	0.948	0.917	0.932
平均	0.954	0.950	0.951

  - 大語彙認識器を除く分別精度：0.978（F 値）
  - タスク依存認識器の単語認識精度：98%

考察

- シンプルで高精度な分類
  - 膨大な学習データが必要ない
- 新たなタスク依存認識器の追加が容易



- 話者ごとのモデルの部分的な切り替えなどが容易
  - 画像による話者認識との統合：マルチモーダル化
- 特定の認識器に依存しない
  - 音素レベルの比較のみで分類が可能
- 統合的な利用に向けて
  - 大語彙認識器の単語認識精度：36%
    - 精度向上が必須

今後の課題

- タスク依存認識器の語彙の拡充
  - メンテナンスツールの開発
- 認識器の統合的利用
  - 大語彙認識器による文脈形成とその有効性の検証
- マルチモーダル化
  - 画像情報との統合：話者認識・発話区間推定