マルチモーダル教師業務ルールベースによる英語教育支援ロボットサービスの開発

阿部秀尚 文教大学情報学部情報システム学科 hidenao@bunkyo.ac.jp

※お断り:ロボットサービスの実装についてはWIP(Work in Progress)の状態です

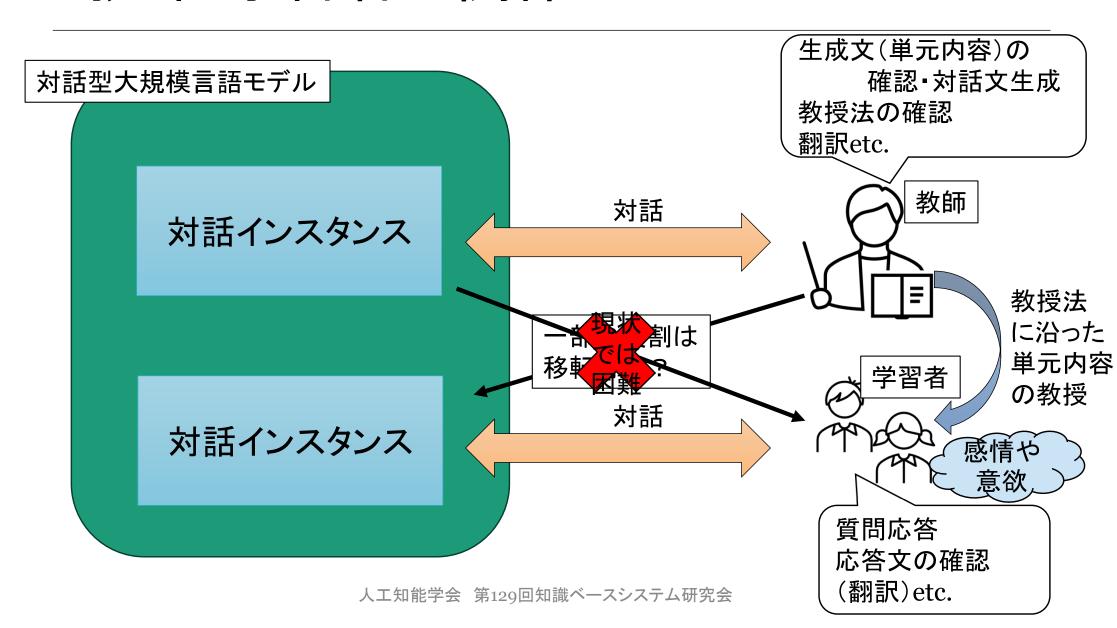
研究背景

- クラウドサービスによるパターン認識・機械学習 アルゴリズム利用環境の発展
 - →画像認識, 音声認識の性能が人の認識と同程度以上に向上
- ・業務知識に基づくロボットサービスの開発による個別の 行動ルール適用知識(メタ知識)の明確化
- ・教師の業務知識を利用した授業支援ロボットサービスの 開発
 - →英語授業支援を目的としたマルチモーダルルールベースに基づくロボットサービスおよび授業支援システムの開発

対話型人工知能(大規模言語モデル) vs. "良い"教師(教育者)

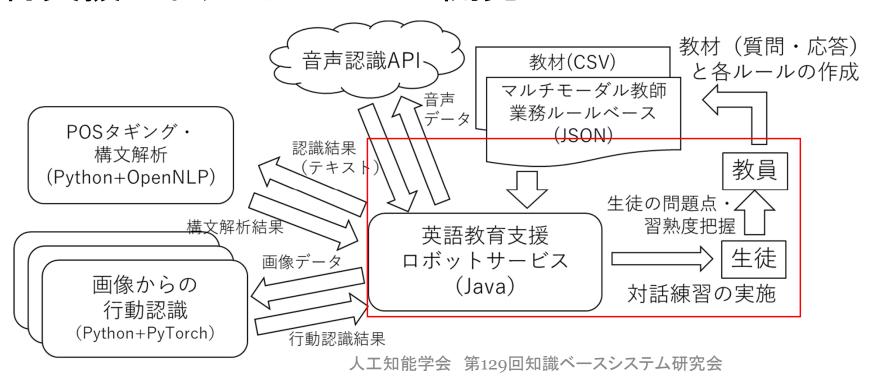
- 大規模言語資源に基づく対話型人工知能
 - GPTモデルに基づくChat GPTなどのマルチタスクな対話が可能に
 - 間違いやでたらめを含む応答文生成
 - →人間が如何に間違いを検出して正す(正しい答えを教え込む)必要
 - 良い質問により、良い答えが導ける(=<u>対話的人工知能と学習者は対等</u>)
- "良い"教師(教育者)
 - 授業では正しい教科知識+教授法に基づく適切かつ臨機応変な対応が必要
 - 適切な対応により学習者の能力を向上(間違いを検出して正し)し, 意欲を引き出す
 - "教え込み" (Instruction)ではなく<u>"教育" (Education)</u>
 - 教授法に基づく適切かつ臨機応変な対応は研究授業や模擬授業を通して 習得
 - →教師(教育者)には<u>メタ知識を伴う技能が求められる(‡学習者の立場)</u>

対話練習における大規模言語モデルと教師・学習者の関係



研究目的

- ・階層的マルチモーダル教師業務ルールベースの記述形式と入力インタフェースの開発→ユーザモデルの記述
- マルチモーダル教師業務ルールベースを利用した 教育支援ロボットサービスの開発



関連研究

- ・ロボットとの質問応答形式による学習効果の評価 [村本2023]
 - ・リピーティング学習(例: Audio-Lingual Method)よりも 質問応答形式による学習で中期的な定着度が向上
- (英語)対話練習におけるICT活用に求められる スキル[Compton2009]
 - ICTそのものを活用するスキル
 - 教授法を扱うためのスキル
 - 教科内容の定着度を評価するスキル

[Compton2009]Compton, L. K. L. (2009). Preparing language teachers to teach language online:

A look at skills, roles, and responsibilities.

Computer Assisted Language Learning, 22, 1, 73–99. doi:10.1080/09588220802613831

先行研究

[秋本2018] 秋本桃子, 阿部秀尚, 生田祐子, 森田武史, 山口高平: 教師業務ルール分析に基づく対話型ロボットを 用いた発音練習の実装と評価, 情報処理学会 情報教育シンポジウム2018 (SSS2018), P-3, 2018. [秋本2019] 秋本桃子, 阿部秀尚, 森田武史, 山口高平:対話型ロボットサービスにおける教師業務ルール実装のための基本動作認識システムの開発, 人工知能学会 第117回知識ベースシステム研究会, 2019.

- ・英語の授業における言語活動での教師の業務ルール(言動パターン)抽出と専門家(英語教職科目担当の大学教員)によるルールの評価
 - →発音矯正指導を行うロボットサービスの実装[秋本2018] により学習者習熟度と業務ルール適用数の差異を確認
- ・音声認識と画像認識による授業環境における基本行動認識性能の評価[秋本2019]
 - →音声認識はクラウドサービスを利用し、基本行動認識では各種 従来型機械学習アルゴリズムによる96%以上の認識精度

先行研究: 学習者習熟度と適用ルールの関連性[秋本2018]

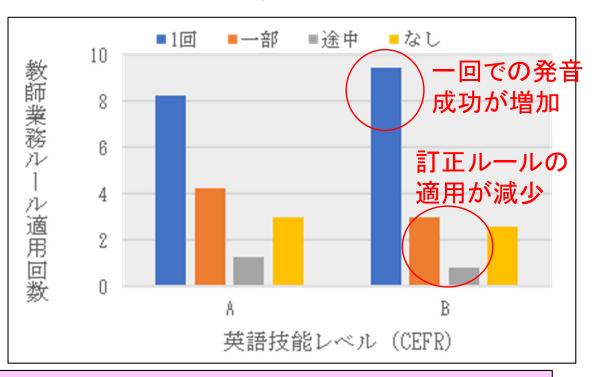
実験内容:

- •Audio-Lingual Methodによる発音練習[教授法]
- •被験者 大学生9名(CEFR A2ランク4名, B1•B2ランク5名)|
- 評価と訂正のためのルール3個(下表)
- ・日本の中学生レベルのランダムな12文

音声認識結果	処理	判定
文頭から8割以上一致	ほめる:Excellent!	一团
文頭から5割以上一致	もう一度発話指示する:One more time.	途中
2割以上一致[誤3単以 内)	英文を発話する:Repeat after me.	一部

実験結果:

学習者の習熟度と 適用されるルール(業務ルール) の関連が示唆される



現場の教師が評価や訂正の行動(ルール)を"適宜"学習者に合わせて適用していることと合致

階層的マルチモーダル業務ルール ベースの構築概観

専門家により言語化された知識 対応付け 専門的に 専門家が理解可能な 言語的記述によるルール

専門家(英語教師)が教授法として言語化している範囲

記号的に 説明可能

パターン認識結果 の組み合わせによるルール 専門家(英語教師)が意識すれば理解できる学習者の様子

信号の 記号化 (認識) 音声認識 自然言語 画像・映像 の認識

個々の技術要素の専門家・研究者により開発される手法による結果

階層的マルチモーダルルールベースの 記述形式

- Droolsの決定テーブルとルール記述形式を基にJSON形式のスキーマを作成
 - 条件部:when, 結論部:then
 - ユーザが用いる言語化された学習者の様子や問いかけ、働きかけに対応する項目を追加

```
"rule": 1つ1つのルール{}

"basic_information": ルールの概要{}

"student_situation": "生徒(学習者)の様子"(自然文)

"example_sentence": "学習者から期待する応答の例文"(自然文)

"parsing_result": "example_sentenceの構文解析結果"

"when": 条件部{}

"word_accuracy": 聞き取りによる対話文の正確さ{}

"voice_recognition": 音声認識結果とその自然言語処理結果[]

"image_recognition": 画像認識による生徒の様子(顔の向きなど)[]

"video_recognition": 複数の画像による生徒の行動認識結果[]

"then": 結論部{}

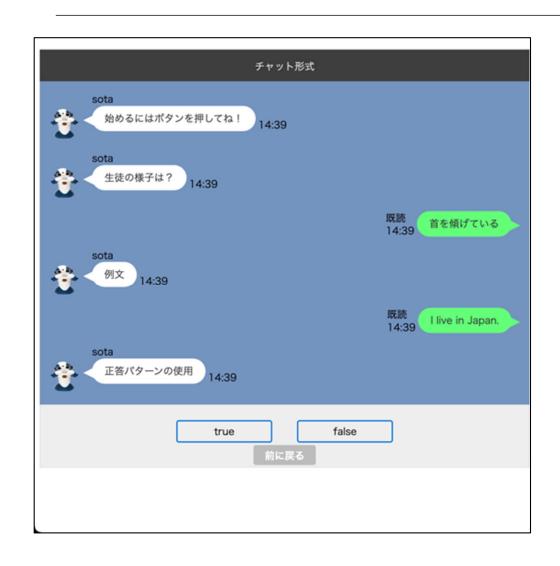
"teacher_behavior": 教師の振る舞い内容{}

"action": ロボットのアクション内容{}
```

階層的マルチモーダルルールベースの 入力インタフェース[阿部2023]

- ・想定する対象ユーザ
 - ・日本語を母語とする英語教員(中・高)
 - 画像認識,音声認識,自然言語処理の処理結果や 処理内容については精通していない
 - 学習者の様子を想定して言語化できる
- 入力インタフェース
 - 2種類のインタフェース
 - HTMLフォームによる入力インタフェース
 - チャット風の入力インタフェース
 - ・使用プログラミング言語
 - インタフェース: HTML/CSS, PHP, Vue.js, Axios
 - 自然言語処理: Python(Flask), OpenNLP

チャット風入力インタフェースによる階層的マルチモーダルルールの入力



- ・テキストや画像, 動画など を含む情報提示を受けなが ら逐次ルールの内容入力
- 想定ユーザが使い慣れている入力インタフェース
 - ・生徒の様子の画像などを 例示しながらの入力が可能
- ・入力した項目の一覧性に 劣る

階層的マルチモーダル教師業務ルールベースに対応したユーザモデル記述形式

- JSON形式によるユーザモデル記述
 - 適用されたルールの履歴
 - 聞き取った学習者からの応答分の正確さの履歴
 - 音声認識結果の履歴
 - 画像認識結果の履歴→行動認識の判定に利用
 - 行動認識結果の履歴
 - 学習者の識別情報

"user_model": ユーザごとのモデル(履歴など){}

"applied_rules": 適用されたルールの履歴[],

"word_accuracy": 聞き取りによる対話文の正確さの履歴[],

"voice_recognition": 音声認識結果とその自然言語処理結果の履歴[],

"image_recognition": 画像認識による生徒の様子(顔の向きなど)[],

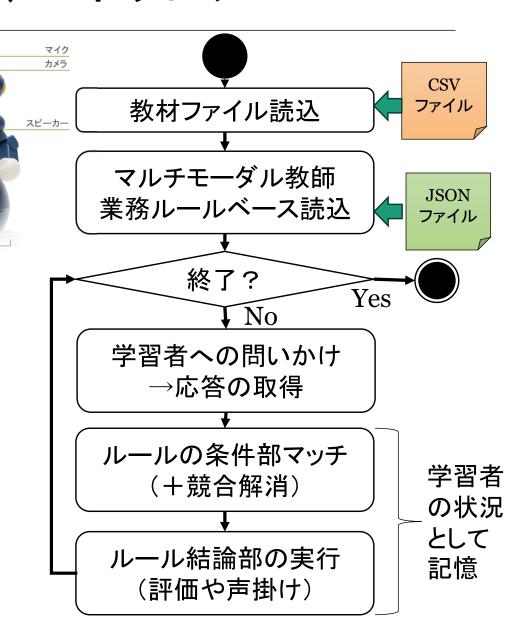
"video_recognition": 複数の画像による生徒の行動認識結果[],

"student_reaction": ルール適用時(action実行時)の学習者の様子[]

マルチモーダル教師業務ルールベースによる英語教育支援(対話練習)ロボットサービス

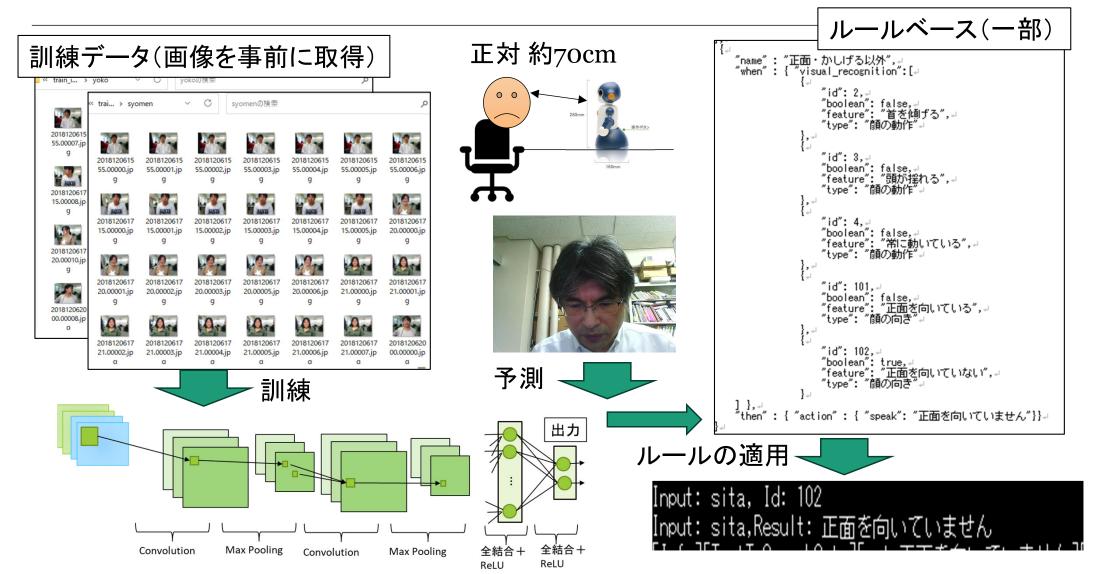
動作環境

- 対話ロボット: VStone社製Sota
 - プログラミング言語:Java
 - マイクやカメラの性能はSotaによる
- 音声認識: IBM Cloud Speech-to-Text
 - 音声認識精度や機能はSTTサービスの品質による
- 自然言語処理: Apache OpenNLP
- 画像認識:独自に作成したCNN
 - プログラミング言語: Python(PyTorch)



人工知能学会 第129回知識ベースシステム研究会

ロボットサービスにおける 画像認識とルールの適用(動作確認)



おわりに

- 英語学習支援を想定した階層的マルチモーダルルール ベースの記述形式とユーザモデル記述形式の策定
- ・階層的マルチモーダルルールベースを利用した英語教育支援(対話練習に基づく学習)ロボットサービスの開発
 - Work In Progress状態
- ・ 今後の課題
 - ・ 階層的マルチモーダルルールベースを利用した英語教育支援(対話練習に基づく学習)ロボットサービスの実装と評価
 - 2形式の入力インタフェースに対するユーザ評価の実施