

階層的マルチモーダル業務ルールベース構築のための インタフェースシステムの開発 -英語教師業務支援を対象として-

阿部 秀尚[†] 池田 啓太^{†‡} 堀内 大希^{†‡}

文教大学情報学部情報システム学科^{†‡}

1. はじめに

近年、深層学習を適用した精度の高い音声認識、画像認識がクラウドサービスを利用可能となってきた。さらに、複雑なニューラルネットワーク構造をもつ深層学習を用いることにより、画像や文章の生成といった複雑なタスクを実行することが可能となっている。しかしながら、データから獲得したパラメータによる関数の組み合わせである複雑なニューラルネットワークでは、その決定過程は人間にとって、ブラックボックスとなる。一方、高精度の認識が可能な各サービスを統合し、人間が聴覚や視覚を駆使して行う業務をロボットサービスとして実現することを行っていくことは、人間の専門家が自身の業務を振り返る上でも重要な洞察を与えることが可能であると考えられる。本研究では、高精度な画像や音声認識、自然言語処理を活用し、人間が行う高度な業務知識の記述や適用を課題とする。

本稿では、教師による英語教育支援を目的とし、音声認識、画像からの姿勢推定、自然言語処理による構文解析の結果などを統合するマルチモーダル業務ルールベースの記述形式の検討と、専門家の業務知識を言語的な表現から個々の認識サービスの認識結果までを階層的に入力するインタフェースの開発について述べる。

2. 階層的マルチモーダル業務ルールベースの構築

画像や音声の認識、およびそれらの結果を処理する自然言語処理といった上位のタスクの精度や性能の向上、およびクラウドサービス化の進展により、ネットワークに接続された機器であれば、組み込み用コンピュータやロボットなどがより多くのタスクを実行することが可能となってきた。この際、業務を実行するための知識は、プログラムとして、その場の適用場面に合わせ作成されているのが現状である。

これに対し、本研究では人間の専門家にとっても理解可能な言語化された業務知識から、図1に示すように機械的に認識が行われるレベルまでを対応させる階層的マルチモーダル業務ルールベースの構築手法を提案してきた[1]。

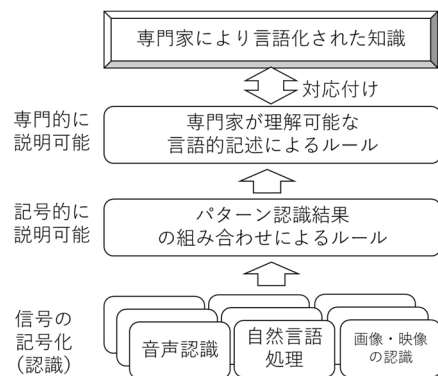


図1: 階層的マルチモーダルルールベース構築アプローチの概要

図1に示すように、音声認識からの自然言語処理による認識結果と画像による行動認識結果を直接、各業務の専門家により言語化された知識と対応付けるのは容易ではない。そのため、階層的に専門家の言語化された業務知識から各認識結果の統合を行うため、本研究では、Drools[2]で用いられるルールベース記述を拡張し、JSON形式による、音声認識および自然言語処理結果、画像による行動認識の結果などを組み合わせるための記述形式を図2に示すように定義した。

```

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

```

図2: 階層的マルチモーダル業務ルールベースの記述形式(JSON形式による)

Development of an Interface System for Acquiring Hierarchical Multi-Modal Business Rulebase for English Teachers

[†] ABE Hidenao, Dept. of Information Systems, Faculty of Information and Communications, Bunkyo University.

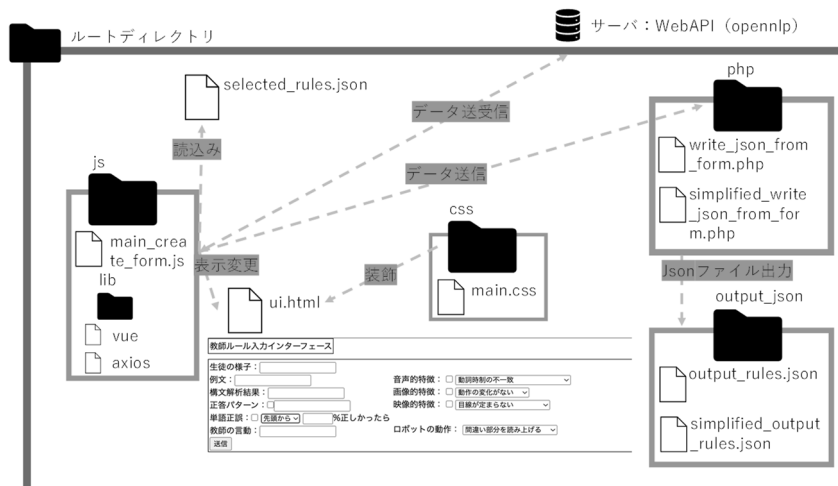


図 3 : 階層的マルチモーダル業務ルールベースの入力システムの概観

3. 階層的マルチモーダル業務ルールベース入力インタフェース

前節で述べた階層的マルチモーダル業務ルールベースの入力を行うため、図3に示す入力インタフェースシステムをWebアプリケーションとして開発した。

本システムでは、今回は中学校以上の英語教育に携わる教師を専門家として想定し、入力を図2に示す階層的マルチモーダル業務ルールベースとしてJSONファイルを出力する。図3では、入力画面にHTMLのformタグで記述可能な表示を用いた表示についても実装を行っている。さらに、図4に示すように、SNSなどで用いられるの対話調のインタフェースの実装も行っている。

4. おわりに

本稿では、音声認識からの自然言語処理による認識結果と画像による行動認識結果を各業務の専門家により言語化された知識と統合した階層的マルチモーダル業務ルールベースを獲得するため、ルールベースの記述形式の検討と入力インタフェースシステムの開発について述べた。

今後は、入力インタフェースについて、SNSなどでの対話調の画面による入力に習熟したユーザに向けた利便性の評価を行っていく予定である。さらに、提案した形式でのルールベースを利用し、英語教育における柔軟な対話練習を実現するロボットサービスの実装を行っていく。

参考文献

- [1] 高橋幸也, 秋本桃子, 阿部秀尚. マルチモーダルルールベースの階層的構築によるクラスルーム教育支援ロボットサービスの開発. 2020年度人工知能学会全国大会(第34回), 4R-in-141, 2020.
- [2] Bali, M.: Drools JBoss Rules 5. X Developer's Guide. Packt Publishing Ltd, 2013.

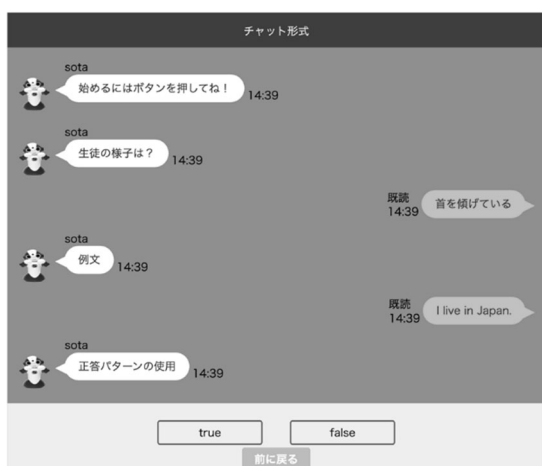


図 4 : 対話調の階層的マルチモーダル業務ルールベース入力画面