

山口 高平

慶應義塾大学工学部
教授

実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践

§ 1. 研究実施体制

(1) 知識グループ

- ① 研究代表者: 山口 高平 (慶應義塾大学工学部、教授)
- ② 研究項目
 - ・ PRINTEPS の開発と評価および多重知識ベース管理システムの研究

(2) 対話グループ

- ① 主たる共同研究者: 中野 有紀子 (成蹊大学工学部、教授)
- ② 研究項目
 - ・ 対話プラットフォームの開発
 - ・ マルチモーダル対話モデルの研究

(3) 画像グループ

- ① 主たる共同研究者: 斎藤 英雄 (慶應義塾大学工学部、教授)
- ② 研究項目
 - ・ 人センシングモジュールの開発
 - 遮蔽された人体の骨格検出における未検出関節の補完
 - 人体遮蔽に頑健な人同定
 - ・ モノ・環境センシングモジュールの開発
 - 物体把持のための三次元物体の基本形状推定
 - 環境の三次元形状の取得とセグメンテーション
 - ・ 音声と画像のマルチモーダル情報を用いたモジュールの開発
 - 音声と画像のマルチモーダル情報を用いた話者特定

- 音声と動画の同期ずれ補正

(4) 対話継続グループ

- ① 主たる共同研究者: 小林 一郎 (お茶の水女子大学基幹研究院、教授)
- ② 研究項目
 - ・ ユーザの興味推定に基づく対話管理モジュールの構築

(5) 知的動作グループ

- ① 主たる共同研究者: 高橋 正樹 (慶應義塾大学理工学部、准教授)
- ② 研究項目
 - ・ 環境における環境認識、知識、音声対話に基づいた行動モジュールの構築
 - ・ ロボット行動モジュールの開発
 - ・ 動作オントロジーに基づく階層型行動計画モジュール
 - ・ 人間行動推定モジュール
 - ・ 人の動きに調和したヒューマンハーモニアス行動モジュール

§ 2. 研究実施の概要

本研究では、エンドユーザが、日本語でワークフロー・業務プロセスを記述すれば、統合知能（知識推論、音声対話、画像センシング、動作計画、機械学習などの要素知能相互作用知能）アプリケーションが自動的に生成される、エンドユーザ向けの統合知能開発プラットフォーム PRINTEPS を開発し、ロボット喫茶店と授業支援ロボットを応用事例として、PRINTEPS の評価を行ってきた。図 1 に研究全体の概要を示す。以下、要素知能ごとに研究実施概要を述べる。



図 1: 研究全体の概要

(1) 知識グループ

平成 29 年度は、PRINTEPS の応用事例としてマルチロボット喫茶店と教師ロボット連携授業の実践を行った。マルチロボット喫茶店については、慶應義塾大学第 18 回矢上祭において、約 100 組の一般客を対象に実践を行った。ロボットサービスにおける顧客満足度を向上させるための重要要因を明らかにするために、マルチロボット喫茶店の来客者にアンケートを実施し、SERVQUAL に基づくサービス品質の評価を行った。その結果、確実性やエンターテインメント性が顧客満足度や再利用意図を促進することが明らかとなった。教師ロボット連携授業については、理科の振り子と運動の単元を対象に、2 つの小学校で各 2 回の実践を行った。これまで開発してきたワークフローエディタを基に、複合イベント処理を利用したルールベースとルール編集インタフェース、ワークフローの再利用機能を新たに実装し、評価を行った。さらに、ロボットやコンテキストに依存しないワークフローを構築できるように、「展開プロセス」と「抽象プロセス」をワークフローエディタに導入すると共に、展開プロセスオントロジーと抽象プロセスオントロジーを用いた展開・具象化ツールの研究開発を行った。ロボット喫茶店における入店時挨拶サービスにおいて、フォーマルな入店時挨拶サービスからカジュアルな入店時挨拶サービスへワークフローを修正する時間を測定し、展開・具象化ツールを用いることにより、156 秒から 23 秒に修正時間を短縮することができた。

(2) 対話グループ

前年度までに開発した対話システムの適用例として、ロボット喫茶店での注文対話を実装した。テーブルに着席している複数名の客に対して、注文を受け付ける対話を行う。図 2 に対話の様子を示す。複数個の注文を同時に行なったり、注文内容を修正、変更することも対話を通して行うことができる。対話によって受け付けた注文内容は、知識モジュールと共有されているデータベースに登録され、その後、ワークフローにしたがって、マニピュレーションロボットに注文が伝えられる。

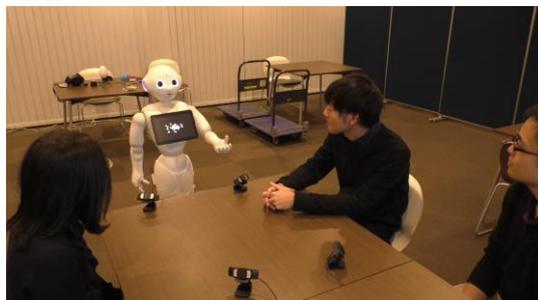


図 2: ロボット喫茶店注文受付対話

マルチモーダルインタラクションの研究においては、①会話中のユーザの優位性や参与役割をリアルタイムに認識し、適切な注視行動を行うことに加え、システム自身に発話権があることを認識することで、適切なタイミングで会話介入を行うことができる対話ロボットを提案した。提案システムにより、優位性の低い参加者の発話を促すことが可能であることが示された。②音声、頭部動作、そして顔特徴点から得られる低次元信号から、議論要約のための会話抜粋を推定するモデルを提案した。本研究では、3次元量み込みニューラルネットワーク(3D-CNN)を用いてモダリティごとにユニモーダルモデルを作成し、さらにそれらを統合してマルチモーダルフュージョンモデルを構築した。研究開始時の人手による特徴量選択を行うモデルでは、F-measure で 0.7 であったが、本モデルでは、F-measure で 0.783 の性能を達成した。

(3) 画像グループ

画像グループでは、実環境における人間動作認識と環境認識モジュールの構築として、人センシングモジュール、モノ・環境センシングモジュール、音声と画像のマルチモーダル情報を用いたモジュール、を開発している。これらは、人の挙動や動作をシステムに認識させるための研究や、カメラにより物体や環境の3次元形状をセンシングし、その認識を行う技術に関連した研究、さらに音声と画像の相互関係等に注目した新しい認識技術の研究を基盤として得られた知見を利用して開発されている。本研究プロジェクトでは、これらのモジュールを、ロボット喫茶店や教育実践といった実際のフィールドでの検証実験等に利用し、本研究プロジェクトの有効性を実証する。

(4) 対話継続グループ

対話における発話の理解および適切な回答を実現するため、ユーザによる発話と大規模知識を想定した Linked Open Data(LOD)で表現した知識とを結びつける自然言語インタフェースを開発した。発話された自然言語文は、構文解析、述語項構造解析を通じて、LOD の RDF トリプルに変換され、発話のタイプ判別、RDF トリプルにおけるプロパティの判別を通じて、知識にアクセスするための SPARQL クエリが生成される。質問のタイプから生成される SPARQL クエリを6種類にタイプ分けをし、事実の問い合わせに回答をする「基本型」、程度に対する問い合わせに回答する「Degree 型」、数に関する問い合わせに回答する「Count 型」、種類の問い合わせに回答する

「List 型」、事実の問い合わせが能動的な表現ではなく受け身的な表現による問い合わせに回答する「受け身型」の6種類について自然言語文から適切な SPARQL クエリを生成することができた。

(5) 知的動作グループ

知能ロボットが安全かつ人の動きに調和してヒューマンハーモニアスに移動するために歩行者の両脚を検出・追跡する技術を開発し、障害物を回避しながら、歩行者を案内する先導走行技術を開発した。障害物回避手法を PRINTEPS 動作モジュールとして実装し、慶應義塾大学矢上キャンパス教員食堂および慶應義塾大学矢上祭(学園祭)で実施したロボット喫茶店において、座席案内および注文物運搬を行う Pepper(ソフトバンク)、HSR(トヨタ)に適用し、複数台のロボットおよび客がいる環境において、5日間全体約24時間の営業時間の中で、衝突することなく安全な移動が実現可能であることを確認した。また、HSRによる複数注文時のカートを利用した運搬や使用済み容器の片づけも動作モジュールとして実装・実践し、アンケートで高い満足度評価を得た。また、HSRなどの移動マニピュレータに対して、障害物回避だけでなくマニピュレーションによりいすなどを動かすことも考慮した階層型行動計画手法を提案した。

- [1] Kodai Nakamura, Takeshi Morita, and Takahira Yamaguchi, "A User-Centric Platform PRINTEPS to Develop Integrated Intelligent Applications and Application to Robot Teahouse", 21st International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2017), Vol. 112, Issue C, pp.2309-2318 (2017) DOI: 10.1016/j.procs.2017.08.266
- [2] 高瀬 裕, 吉野 堯, 中野 有紀子: 多人数会話における調整・介入機能を有する対話ロボット, 情報処理学会論文誌, Vol. 58, No. 5, pp. 967-980, 2017
- [3] T. Kikuchi and Y. Ozasa, "WATCH, LISTEN ONCE, AND SYNC: AUDIO-VISUAL SYNCHRONIZATION WITH MULTI-MODAL REGRESSION CNN", in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018. IEEE, 2018, pp. 3036-3040