

マルチモーダルインタラクションを目指したオントロジーロボット

Ontology Robot aimed at Multi-Modal Interaction

樋川 暁 森 雄一郎 小川 雄平 山口 高平
Akatsuki HIKAWA Yuichiro MORI Shinpei OGAWA Takahira YAMAGUCHI

慶應義塾大学
Keio University

We build various HRI(Human-Robot Interaction) where humanoid robot can make dialogue and actions well with users and multi-robot cooperate, aligning different kinds of ontologies: Japanese Wikipedia Ontology(JWO), Behavior Ontology and Robot Kinematic Ontology. In this paper, We focus on visual information contained in human behaviors and aim to realize multi-modal HRI combined verbal and non-verbal knowledge.

1. はじめに

我々はコミュニケーションロボットの大部分が言葉の意味を理解することが出来ないことが、HRI(Human-Robot Interaction)が上手くいかない原因であるという観点から、これまでに、本研究室で日本語 Wikipedia から半自動的に構築した大規模汎用オントロジーである日本語 Wikipedia オントロジー[玉川 11]を対話知識として利用し、ロボットの実行可能な振舞いについて、HRIに特化した領域オントロジーを手動構築し、関連付けすることにより、言葉の世界、振舞いの世界、組み込みソフトウェアの3つの異なる世界をインテグレーションし、対話と動作の連携を実現するシステムを提案してきた。[Kobayashi 11]

さらに、生活環境におけるトータルな HRI の実現を目指し、異なる身体的特徴を持つ異機能ロボットの身体性に特化した部位オントロジーとして手動構築し、既存のオントロジーと連携することにより、環境やタスクに適したロボットが処理を分担する異機種ロボット連携 HRI システムを提案した。[小林 12]

しかしながら、これまで提案してきたシステムでは、インタラクションが発生する起点が一つしか存在せず、ロボットがユーザに話しかけ、ユーザの要求に対してロボットがタスクを実行する状況に限られている。生活環境においてロボットが積極的にユーザにサービスを提供するためには、聴覚による言語情報の利用に加えて、ユーザの振舞いに含まれる非言語情報の利用が重要であると考えた。本稿では新たに非言語情報の利用に基づく HRI を実現する。具体的にはロボットが作業中のユーザを監視し、ユーザの振舞いに含まれる非言語情報について、対話を通じて原因を特定し、特定結果に対応したサービスを提案するシステムを実現する。システムの有用性について、ケーススタディを通じて示す。

2. 関連研究

オントロジーは、これまでもロボティクス分野の研究で利用されている。Johnston らは、OBOC (Ontology Based Object Categorization) [Johnston 08]を開発し、オントロジーのシンボルグラウンディング問題への利用を提案し、その有効性をロボカップで検証している。また、Hong Suh らは、オントロジーとルールによりロボットの知識フレームワーク OMRKF (Ontology-based Multi-layered Robot Knowledge Framework) [Suh 07]を構築し、

連絡先: 樋川暁, 山口高平, 慶應義塾大学理工学部
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
TEL:045-566-1614
akatsuki@a2.keio.jp, yamaguti@ae.keio.ac.jp

その有用性を運搬タスクによって示している。しかし、オントロジーに基づき、Wikipedia の膨大な情報を利用してユーザと対話を行い、さらに、動作レベルでもユーザとの双方向コミュニケーションを実現する研究はなされていない。本稿では、動作は作りこみであっても大規模になればよりよい HRI を実現できるという観点から、ロボットの動作を動作オントロジーとして事前に体系化した。

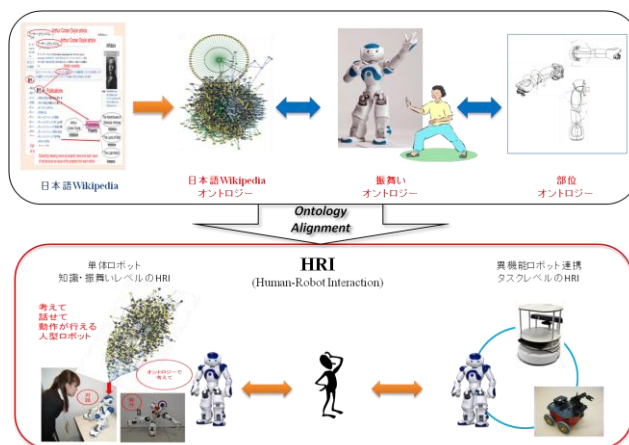
ロボットインタラクションにおける非言語情報を利用した研究もこれまでにに行われている。しかし、既存研究では対話中にロボットがうなずきや相づちの行動を示す、ユーザの表情や視線を観察するなどあるが、これらは自然な対話の実現や対話時間の延長を目的としている。本研究のように対話を行っていない時にユーザの示す振舞いに注目し、インタラクションに利用する目的としては利用されていない。

3. システム概要

3.1 既存システム概要

はじめに、我々がこれまでに構築してきたシステムの概要を図1に示す。

図1 既存システム概要



我々の研究目標は異なる世界の知識を表現するオントロジーを連携させ、意味理解に基づく HRI のサービスを提供することである。これまでに知識用語の世界である日本語 Wikipedia オントロジーと HRI に特化した振舞いオントロジーを関連付け

することにより、人型ロボット Nao を利用した対話と動作の融合による HRI を実現した。さらにロボットの持つ身体的特徴を部位オントロジーとして構築、振舞いオントロジーに関連付けすることにより、タスク分担処理によるトータルな HRI を実現した。

3.2 提案システム

本稿では、ユーザの振舞いを認識した際に、ユーザの振舞いの原因を推定し、ユーザの潜在的なニーズに対するサービスをロボットが提案することを目指す。

しかし、ユーザが行う振舞いは無数に存在する。さらに一つの振舞いに対しても、ユーザが存在する環境に応じて、振舞いが発生する原因は膨大である。したがって、全ての原因の中から推定を行い、対応することは困難である。本稿では、ユーザが存在する環境の状況と認識する振舞いを限定した条件のもと、ロボットがユーザの振舞いを認識し、与えられた候補の中から対話を通じて原因を特定しサービスを提供するシナリオを実現する。

以下図 2 に本稿のシステム概要図を示す。

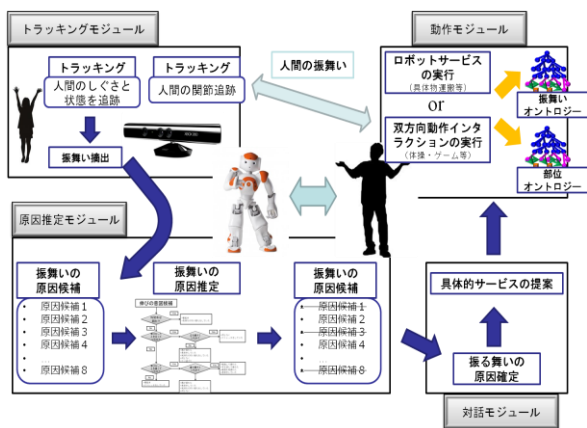


図 2 本稿システム概要図

図2に示したように、提案システムは大きく4つのモジュールから構成されている。特に本稿ではトラッキングモジュールと原因推定モジュールについて説明する。

3.3 トラッキングモジュール

トラッキングモジュールではユーザの各関節角度を取得しユーザの姿勢を決定する。関節角度の取得を行うタイミングは、以下の2つの時点である。

- Nao がユーザを監視している時
- 双方向動作インタラクションの実行時

なお、ユーザの関節角度を取得するにあたり、本研究では Microsoft 社製 Kinect センサーを用いた。Kinect は Xbox 360 用のゲーム用のセンサーであるが安価であり、公式 SDK が無料で公開されているなど、開発が容易である点から多くの研究でも利用されている。

本研究では2つの時点における、関節角度の利用方法が異なるため、2つのトラッキングプログラムの実装を行った。前者のトラッキングではユーザが任意のタイミングで振舞いを実行するため、ユーザの全身の関節角度に注目し、監視中の全ての振舞いと状態(立つ、座る)を保存し続けている。ユーザがある時点で

特定の振舞いを示した場合はその振舞いと状態を抽出し、原因推定モジュールへ展開する。一方後者のトラッキングは動作モジュールの双方向動作インタラクションが展開されている際に実行され、Nao が振舞いオントロジーのインスタンスに記述されている振舞いを実行し、ユーザがそれに合わせて動作を行うため、特定の関節についてのみ関節角度の取得を行い、ユーザの動作状況(各関節角度の時系列データ)を判定する。Nao が提案した動作インタラクションの内容(振舞いインスタンスのクラス)により、実行中のインタラクション(振舞いの実行順序や難易度)を動的に変更することが出来る。例えば Nao が提案したサービスが体操だった場合は判定結果に基づき、以下のように動作インタラクションを動的に変更することが可能である。

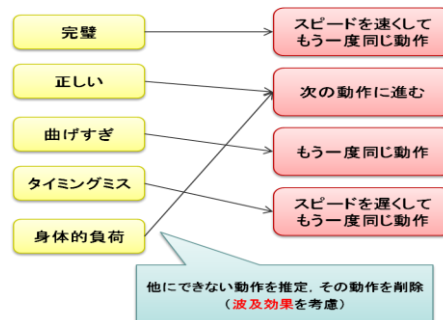


図 3 判定結果とプラン決定

また、提案した動作インタラクションが簡単なゲームのような場合では時系列データに基づく判定結果をそのままユーザへ返す。

3.4 原因推定モジュール

原因推定モジュールは、抽出された振舞い、ユーザの状態、外部環境に基づき、振舞いの原因を推定する。振舞いの原因とは、意図を特定することである。オントロジーの視点から考えると意図には様々な世界の事象が含まれている。病気に関する世界や生活環境に関する世界、心理に関する世界など、非常に多くの世界を1つのオントロジーとして定義することは困難であるが、それら全てを網羅することが本研究における目的ではないため、振舞いの原因が関連付けされていると仮定した上で本研究を行った。手動構築した原因候補の例を図5に示す。なお、抽出可能なユーザの振舞いは振舞いオントロジーのインスタンスとして定義されている。

振舞いの原因候補から条件分岐によってさらに候補を絞り込んでいく。原因候補が設計者の主観によってきめられているため、分岐による絞り込みだけでは原因を一意に定まらないように設定した。絞り込まれた原因候補は Nao がユーザと対話を行うことにより、決定する。推定した結果が正しいとユーザに判定された場合は、Nao は振舞いの原因を特定できたと判断し、具体的なサービスの提案を行う。振舞いの原因と提案するサービスの一例を図4に示す。推定された候補とユーザの判定結果が一致しない場合は条件分岐を戻り、候補を拡張した後、対話を行うことによって決定を行うが、ユーザの負担を考慮し、さかのぼる過程は2回までとした。Nao が提示した原因候補とユーザの判定結果が一致しない場合には、既存システムに従い、Nao がユーザから音声によるタスク要求を促す。

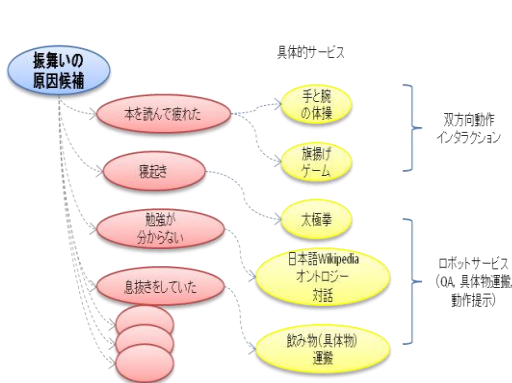


図4 サービス提案例

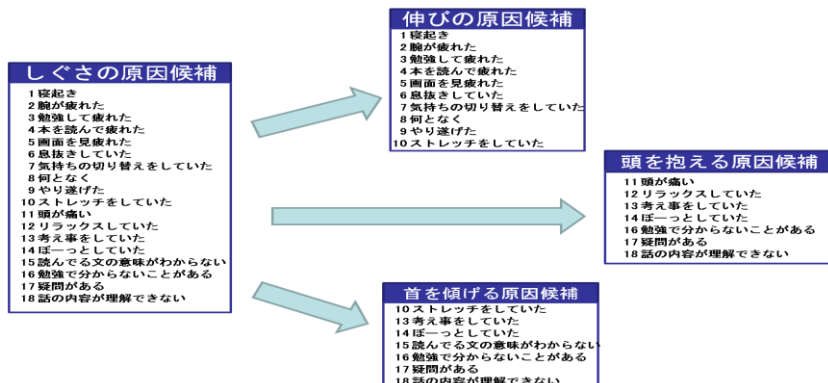


図5 振舞いの原因候補例

4. ケーススタディ

4.1 実験環境

本稿での実験環境およびシナリオについて説明する。実験環境は本研究室内とし、ユーザの関節角度を正しく取得するため、Kinect の撮影可能範囲内にはユーザのみが存在する状況下で実験を行った。

4.2 ケーススタディ

ケーススタディではユーザは本を読んでおり、Kinect を通じて Nao がユーザの振舞いを認識し、推定と対話を通じて振舞いの原因を特定し、旗立てゲームの動作インタラクションへ展開していくシナリオの流れを示す。旗立てゲームの判定は時系列データに基づき Nao よりもユーザが早く正しい姿勢をとれば、ユーザの勝ち、遅ければ負けとし、5回実施した時の勝利数によって勝敗を決定する。なお、(U)はユーザ、(N)は人型ロボット Nao の発言を示す。

<<トラッキングによる振舞い抽出>>

[[人間が座っている]]

<< Kinect により、ユーザーの状態を認識(図 6)>>

[[人間が伸びをする]]

<< Kinect により、ユーザーの振舞い“伸び”を認識(図 7)>>



図6 ユーザの状態を認識

図7 振舞い“伸び”を認識

<<伸びをした原因候補から、条件分岐により原因推定を行い、候補を絞り込む>>

<<絞り込んだ結果を基に、対話を行い、原因の特定を行う(図 8)>>

N 「勉強して疲れたから、伸びをしていたの？それとも、本を読んで疲れたから伸びをしていたの？」

U 「違うよ。」



図8 原因推定に基づく対話

N 「じゃあ、本を読んで疲れたから、伸びをしていたの？それとも、息抜きをしていたから伸びをしていたの？」
U 「本を読んで疲れたから、伸びをしたんだよ。」

<<“伸び”をした原因を特定、サービスの提案>>

N 「本を読んで疲れたから、伸びをしていたんだね。」
N 「本を読んで疲れたのなら、気分転換と一緒に僕とゲームする？」

U 「うん。」

N 「じゃあ、僕と一緒に旗上げゲームをしよう。」

<<旗上げゲームへの展開(図 9)>>



図9 旗上げゲームへの展開

N 「これからゲームの説明をするね。全部で5回ゲームをするよ。ゲームが始まる前は、僕と同じポーズをとってね。僕が話し始めたら、僕より早くポーズをとってね。赤なら右手、白なら左手を動かしてね。」

U 「うん。」

<<ゲーム開始>> 以下のサイクルを5回繰り返す。なお上げる旗とセリフは5回ともランダムで決まる。

<<1サイクル開始>>

N 「右上げて、左下げない」 ※ここは毎回異なる

<<トラッキングによる関節角度の時系列データ取得>>

<<正しい姿勢を認識したら、実行時間の判定を行う>>

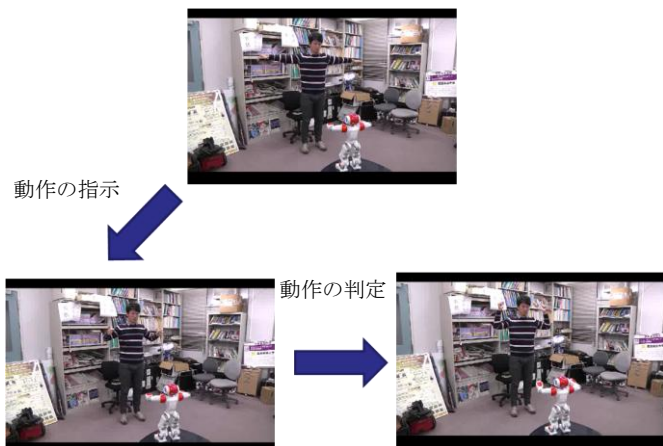


図 10 旗上げゲームの実行例

N 「次の動作に移るね。」
 <<1サイクル終了:5回ゲームを行う>>
 <<ゲーム終了. 勝敗を判定. >>

N 「このゲームはあなた(ユーザ)の勝ちです。」
 N 「これで終わりだよ. 気分転換出来た?」
 U 「ありがとう。」

上述のケーススタディでは Nao がユーザの振舞いを認識し、推定と対話を通じて、非言語情報である振舞いの原因を特定し、よりユーザにとって価値が高いと考えられる具体的なサービスへインタラクションを展開することが出来ることを示している。

5. おわりに

本稿では、ユーザの振舞いに含まれる非言語情報を取得し、原因を推定することによって、ロボットが積極的にユーザにサービスを提案し、ユーザの状態にあったインタラクションを実現することが出来た。人間のニーズを言語情報以外の情報から取得することによって、ロボットがユーザの行動に関して気づきを示すことができ、既存システムの課題であったインタラクションの起点が一方的であることを解消することが出来た。また、ケーススタディを通じて本研究の有用性を示した。

しかしながら本研究はまだ多くの課題を残している。本研究では認識可能な振舞いの数が非常に少なく、ユーザの示す振舞いの多くが見落とされている。また振舞いの推定においても、原因候補や条件分岐の多くに主観的な要素が含まれており、必ずしも適切な原因推定を行うことが出来ていない。3.4 でも述べたとおり外部環境や心理状態の変化によって振舞いを行う意図は無限に変化する。それらの全てを網羅することは非常に困難であるが、本研究では既存システムで構築されたロボットとユーザの対話を通じて原因推定の精度を向上させるように努めた。

HRI において、ユーザが予めロボットの実現可能なサービスを認識しておくことは現実的な方法ではないため、本研究のようなユーザの状態に応じたサービスをユーザが必要とするタイミングで実現することがよりインタラクションの質を向上させる上では重要な要素であると考えられる。

今後の展望として抽出可能な振舞いの数の増加、外部環境や時間帯、ユーザの事前状態の把握、振舞いに影響を与える要素情報の体系化を行い、さらに異なる非言語情報の利用を考え、インタラクションの高度化を目指すつもりである。

参考文献

- [玉川 11] 玉川 奨, 森田 武史, 山口 高平, "日本語 Wikipedia からプロパティを備えたオントロジーの構築", 人工知能学会論文誌 特集論文「近未来チャレンジ」 Vol. 26 No. 4 pp. 504-517 (2011)
- [小林 12] 小林 昭太郎, 樋川 暁, 山口 高平, "対話と動作の連携を目指したオントロジーに基づく知能ロボット", 人工知能学会全国大会(2012), 3K1-R-11-7
- [Kobayashi 11] Shotaro Kobayashi, Susumu Tamagawa, Takeshi Morita and Takahira Yamaguchi: Intelligent Humanoid Robot with Japanese Wikipedia Ontology and Robot Action Ontology, HRI2011 (6th International Conference on Human-Robot Interaction), pp. 417-424 (2011)
- [Winograd 72] Winograd, Terry: Understanding Natural Language, Academic Press, (1972). (Revised version of Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. AI-TR-17, Cambridge, Mass: MIT AI Lab., 1971.)
- [Johnston 08] Benjamin Johnston, Fangkai Yang, Rogan Mendoza, Xiaoping Chen, Mary-Anne Williams: Ontology Based Object Categorization for Robots, Proceedings of the 7th International Conference, Practical Aspects of Knowledge Management, pp. 219-231 (2008).
- [Suh 07] Il Hong Suh, Gi Hyun Lim, Wonil Hwang, Hyongwon Suh, Jung-Hwa Choi: Ontology-based Multi-layered Robot Knowledge Framework (OMRKF) for Robot Intelligence, Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 429-436 (2007).