

# 居住空間において生活者の日常生活を見守り支援する ロボットの基盤技術開発

Development of a Basic Technology of Watching for Residents' Activity at the Residential Environment

2151014



研究代表者 お茶の水女子大学 基幹研究院 教授 小林 一郎

## [研究の目的]

高齢化社会の到来により、近い将来、家庭にロボットが導入され、ロボットは居住者の生活を見守り、居住環境の管理を担う生活のパートナーになることが考えられる。ロボットは居住環境及び居住者の状態を常に観察・認識し、居住者とのコミュニケーションに反映させ、必要に応じて第三者に居住者の状況を説明する役割を担うと考える。このことから、本研究においては、ロボットと居住者のマルチモーダル対話を実現し、ロボットがセンサから観測した居住者の様子や言動を言語で表現し、居住者のプライバシーを保護しつつ第三者に伝達できる見守りシステムの基盤技術の構築を行う。

## [研究の内容, 成果]

### 1. ロボットとのマルチモーダル対話

#### 1.1 マルチモーダル情報の観測

ロボットはソフトバンクロボティクス社とアルデバランロボティクス社が共同開発した、感情認識ヒューマノイドロボット Pepper (<http://www.softbank.jp/robot/>) を使用する。Pepper の様々なセンサからマルチモーダル情報を取得し、それに基づくコミュニケーションを実現する。具体的には、マイクから音声情報、RGB カメラから表情などの画像情報、タッチ

センサから触覚情報、レーザーセンサやソナーセンサから距離情報を取得する。画像情報を用いた顔認識では、ユーザに対して、個体の識別、年齢の推定、笑顔度の判定、5種類の表情 {無表情, 幸せ, 驚き, 怒り, 悲しみ} の識別を行うことができる。

#### 1.2 部分観測マルコフ決定過程

本研究では、実環境での観測情報の不確実性を考慮するため、部分観測マルコフ決定過程 (POMDP: Partially Observable Markov Decision Process) の枠組みを用いる。一般的に POMDP の観測状態は、 $\{S, A, T, O, Z, R, b\}$  で表される。 $s \in S$  はユーザ状態、 $a \in A$  はシステムの行動を表す。また、 $T$  は行動  $a$  によって状態  $s$  が  $s'$  へと遷移する確率 (状態遷移確率  $P(s'|s, a)$ ) の集合であり、 $o \in O$  はユーザから観測される観測値を表す。 $Z$  は行動  $a$  によって状態  $s'$  に遷移し、観測値  $o'$  が観測される確率 (観測値出力確率  $P(o'|s', a)$ ) の集合である。 $r(s, a) \in R$  は状態  $s$  で行動  $a$  を行った時の報酬を表す。POMDP では、観測値  $o$  から直接観測できない状態  $s$  を確率分布として推測し、その分布を信念状態  $b(s)$  とする。初期信念状態を  $b_0$  と表す。信念状態  $b(s)$  が既知のとき、状態遷移確率と観測値出力確率により、次の時刻の信念状態  $b'(s')$  は以下の漸化式で記述される。

$$b'(s') = k \cdot P(o'|s', a) \sum_s P(s'|s, a) b(s)$$

ここで、係数  $k$  は  $\sum_s b'(s) = 1$  を満たす正規化項である。

### 1.3 マルチモーダル状態表現への拡張

ユーザとのインタラクションを想定して、下記に示す3つのユーザ状態  $s^e, s^p, s^l$  が考えられる。

・心理状態： $s^e$

喜怒哀楽のようなユーザの心理的な状態を示す。画像情報を用いた表情認識を用いて観測  $o^e$  を取得する。

・物理状態： $s^p$

ユーザがロボットからどれくらいの距離にいるのか、ロボットに触っているかないかなどの物理的な状態を示す。観測  $o^p$  はレーザーセンサやソナーセンサ、タッチセンサから取得する。

・言語による情報交換： $s^l$

「おはよう」などの挨拶や、「～してほしい」という要求のような、ユーザの発話による情報交換を示す。観測  $o^l$  は音声情報から取得する。

この3つの状態に対応する観測をそれぞれ  $o^e, o^p, o^l$  とする。

### 1.4 状態の階層的關係

強化学習において、状態空間の増大によって最適方策を求めることが困難な場合、階層的強化学習を適用することで解決することができる [Kishimoto 2010]。階層的強化学習では、複雑なタスクを階層的に分解し、各部分問題に対して局所的政策を学習したのち、それらを統合することによって大域的な政策を学習する。本研究ではマルチモーダル状態に対応させて、心理タスク、物理タスク、言語タスクに分割する。

表1にマルチモーダル対話例と、図1に実際の対話の様子を示す。また、階層型の POMDP をグラフィカルモデルで表したものを図2に示す。

### 2. 1日の報告文の生成

次に、Pepper が観測した内容を時系列データとしてデータベースに記録し、1日分のデー

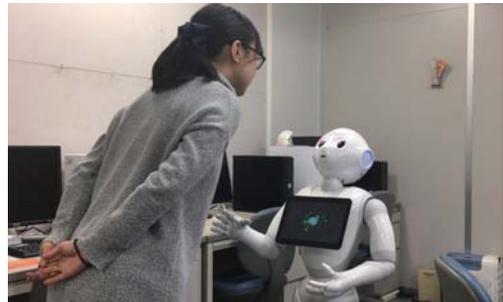


図1 実際の対話の様子

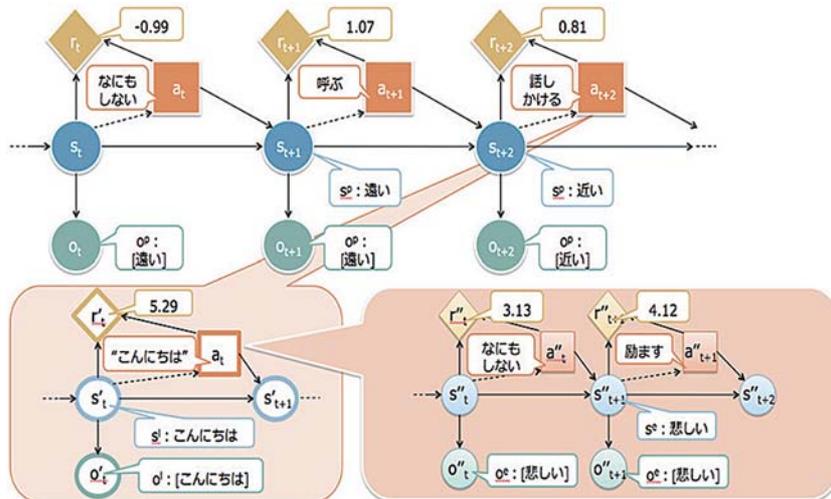


図2 階層型 POMDP のグラフィカルモデル

表1 対話例

話者	発話・行動	観測	b(s)
ユーザ Pepper	(遠くにいる) 近くにおいでよ	$o^p$ (遠い)	
ユーザ Pepper	(近くにくる) 僕とお話しよう	$o^p$ (近い)	
ユーザ Pepper	こんにちは こんにちは	$o^l$ (こんにちは)	
ユーザ Pepper	(暗い顔) 疲れた顔をしてるね	$o^e$ (悲しい)	

タに対する報告文の生成を行う。

## 2.1 Pepper による情報の取得

Pepper から観測した情報をデータベースに格納する。今回は記録する情報として、{日付, 時間, ユーザ名, 距離情報, 表情情報, ユーザの発話内容, ロボットの発話内容} と設定する。ここで、ユーザ名とは Pepper が対話、もしくは観測した一人の相手のことを指す。距離情報は Pepper からの距離を記録し、表情情報は、5 種類の表情 {無表情, 幸せ, 驚き, 怒り, 悲しみ} を数値で記録する。また、ユーザ名, 距離情報, 表情情報, ユーザの発話内容, ロボットの発話内容については観測状況によっては取得できないことがある。その場合、取得できなかった情報は“-”として記録する。ユーザ名, 距離情報, 表情情報, ユーザの発話内容, ロボットの発話内容のすべてのデータが観測できなかった場合には観測がなかったとしてデータを記録しないこととした。また、ユーザの発話内容に対するロボットの発話等の対話管理には POMDP を使用した。

## 2.2 重要情報の抽出

あるデータにおいてひとつ前のデータと比較し、変化のない場合にはそのデータを削除することにより、データベースの冗長性とデータサイズを削減する。データベースへ格納した時系列データに重要度を付与することで、重要な情報を効率的に伝えることができるようになる。ここで、今回のデータにおける重要度について考える。本提案では以下の点に該当するデータ

を重要だと判断する。

- ・ユーザ名, 距離情報, 表情情報, ユーザの発話内容, ロボットの発話内容ができるだけ多く観測できたデータ
  - ・観測対象 (ユーザ) が変わったデータ
  - ・ユーザの表情情報が一時刻前より大きく変動したデータ
  - ・ユーザの発話内容に入力があったデータ
  - ・ロボットの発話内容に入力があったデータ
- これらの条件のいずれか、または全てを満たすデータに重要度を付与する。

また、観測の場面の変化の記録のためにデータを時間で区切ることを考える。時間で区切り、その時間内での重要情報をだけを残すことで、データ量を削減しつつ情報の取りこぼしを減らす。時間によって区切る方法としては、ある時刻に観測したデータの時間とその一時刻前に観測したデータの時間を比べて、その差が5分以上あった場合には場面が変わっていると区切ることにする。

## 2.3 発話内容の特定

ユーザとロボットとの発話の内容の特定を行う。今回は会話の内容を2つのトピック {天気, ご飯} に分類する。トピックの分類のため、Wikipedia における「天気」と「飯」の概要と、時間で区切った範囲のユーザとロボットの対話内容を単語分割し、文書ベクトルを作成する。それらの文書ベクトルに対して cos 類似度を測ることで、ユーザとロボットの対話内容に最も近いトピックを推定する。

## 2.4 テンプレートによる文生成

重要度に対して適切な閾値を設定し、その閾値よりも値の高い文を選ぶことによって重要な情報を取り出し、その情報にテンプレートを適用することにより要約文を生成する。「“いつ”, “だれ” と会いました。その人は “どんな様子” でした。」もしくは「“いつ”, “だれ” と会いました。その人と “何の” 話をしました。」とい

うテンプレートに情報を入れて文を生成する。「“どんな様子” だったか」については表情情報において最も数値の大きい値を使用し、その値と他の値との差分を計算し、差分が大きい場合には「とても」、小さい場合には「少し」という副詞をつける。実際に Pepper の実機を動かして、取得したデータについて文生成を行った。「“何の” 話をしたか」というテンプレートは観測したデータのユーザの発話内容とロボットの発話内容に入力があつた場合に使用する。表2に示したデータから表3に示す重要なデータを取り出し、そのデータに対して文を生成する。重要データ {時間, ユーザ名, 距離情報, 表情情報, ユーザの発話内容, ロボットの発話内容}={12: 20, iijima, 0.49, {0.34, 0.0, 0.58, 0.0, 0.08}, 今日は寒いね, 今日の最高気温は10度だって}, {12: 21, iijima, 0.47, {0.81, 0.0, 0.02, 0.0, 0.17}, 明日は晴れるのかな, 明日の天気は晴のち曇みただよ}では、ユーザの発話内容とロボットの発話の内容に入力があるので、「“い

つ”, “だれ” と会いました。その人と “何の” 話をしました。」というテンプレートを使用する。ここでユーザとロボット対話内容は『今日は寒いね 今日の最高気温は10度だって 明日は晴れるのかな 明日の天気は晴のち曇みただよ』となる。これを単語分割して文書ベクトルを作成する。この文書ベクトルと「天気」についての文書ベクトル, 「飯」についての文書ベクトルについて cos 類似度を測る。「天気」との cos 類似度は0.11, 「飯」との cos 類似度は0.0となるので、今回のユーザとロボットの発話内容は「天気」についてであると推測する。よって、「12時20分ごろに飯島さんと会いました。飯島さんと天気についてお話をしました。」となる。また、重要データが {14: 55, iijima, 0.69, {0.02, 0.10, 0.51, 0.05, 0.32}, -, -} の時には「“いつ”, “だれ” と会いました。その人は “どんな様子” でした。」というテンプレートを使用し、「14時55分にも飯島さんと会いました。少し驚いたような様子でした。」

表2 取得データ抜粋

時間	ユーザ名	距離	表情 {無表情, 幸せ, 驚き, 怒り, 悲しみ}	ユーザの発話内容	ロボットの発話内容
12: 20	iijima	0.49	{0.34, 0.0, 0.58, 0.0, 0.08}	今日は寒いね	
12: 20					今日の最高気温は10度だって
12: 21	iijima	0.47	{0.81, 0.0, 0.02, 0.0, 0.17}	明日は晴れるのかな	
12: 21					明日の天気は晴のち曇みただよ
12: 22	iijima	0.50	{0.66, 0.0, 0.15, 0.04, 0.15}	-	-
12: 22	-	1.12	-	-	-
12: 23	-	1.12	-	-	-
12: 23	-	1.12	-	-	-
12: 24	-	0.96	{0.0, 0.0, 0.01, 0.0, 0.99}	-	-
12: 25	iijima	0.73	{0.12, 0.0, 0.51, 0.05, 0.32}	-	-
14: 54	-	1.17	-	-	-
14: 54	-	0.82	{0.0, 0.0, 0.11, 0.10, 0.79}	-	-
14: 55	iijima	0.69	{0.02, 0.10, 0.51, 0.05, 0.32}	-	-

表3 重要データ抜粋

時間	ユーザ名	距離	表情 {無表情, 幸せ, 驚き, 怒り, 悲しみ}	ユーザの発話内容	ロボットの発話内容
12: 20	iijima	0.49	{0.34, 0.0, 0.58, 0.0, 0.08}	今日は寒いね	
12: 20					今日の最高気温は10度だって
12: 21	iijima	0.47	{0.81, 0.0, 0.02, 0.0, 0.17}	明日は晴れるのかな	
12: 21					明日の天気は晴のち曇みただよ
14: 55	iijima	0.69	{0.02, 0.10, 0.51, 0.05, 0.32}	-	-

[今後の研究の方向, 課題]

本研究では, Pepper を対象にした見守りシステムの構築を目的とし, マルチモーダル対話を POMDP の枠組みに沿って実装を行った。また, 対話状態を含んだマルチモーダル情報に基づいた要約文生成についても実装を行った。今後の課題として, 生成した文に対して評価を行うことが挙げられる。

[成果の発表, 論文等]

- 飯島采永, 小林一郎: マルチモーダル情報を対象とする対話処理への一考察, 言語処理学会第 22 回年次大会発表論文集, pp. 7-9, 2016
- 飯島采永, 小林一郎: POMDP を用いたマルチモーダル対話への取り組み, 第 78 回情報処理学会全国大

- 会発表論文集, 4-213-214, 2016
- 飯島采永, 小林一郎: ロボットを用いたマルチモーダル対話実現への取り組み, 第 30 回人工知能学会全国大会発表論文集, 203-1in1, 2016
- Sae Iijima, Ichiro Kobayashi: A POMDP-based Multimodal Interaction System Using a Humanoid Robot, The 30th Pacific Asia Conference on Language, Information and Computation, 2016
- 飯島采永, 小林一郎: マルチモーダル情報に基づく見守りシステム開発への取り組み, 言語処理学会第 23 回年次大会発表論文集, pp. 470-472, 2017
- 飯島采永, 小林一郎: マルチモーダル対話を用いた見守りシステム構築への取り組み, 第 79 回情報処理学会全国大会発表論文集, 5Q-06, 2017
- 飯島采永, 小林一郎: マルチモーダル情報を用いたロボットによる見守りシステム開発の一考察, 第 31 回人工知能学会全国大会発表論文集, 1D1-2in2, 2017