

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

A Study on Development of a Nursing Care System by Robot using Multimodal Information

飯島采永^{*1}

Sae Iijima

小林一郎^{*2}

Ichiro Kobayashi

^{*1}お茶の水女子大学大学院 人間文化創成科学研究科

Department of Information Sciences, Faculty of Science, Ochanomizu University

^{*2}お茶の水女子大学 基幹研究院 自然科学系

Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

In the future, home robots are expected to spread to homes. The robot observes multimodal information and realizes communication with people. In this research, we will develop a system to grasp the behavior of people from logs of such communication between robots and people and to watch over them.

1. はじめに

近年、家庭用ロボットが多く普及されてきている。ロボットと共に生活していく上で、ロボットのコミュニケーション能力のさらなる充実が、今後、益々必要と考えられる。そこで、本研究では家庭用ロボットの身体性を利用し、生活を見守るためのシステムの構築を目的とし、その実現に向けて、マルチモーダル情報を用いた部分観測マルコフ決定過程に基づいたロボットとの対話処理と、ロボットの観測に基づく1日の要約文やそれに基づく質問応答の作成に取り組む。

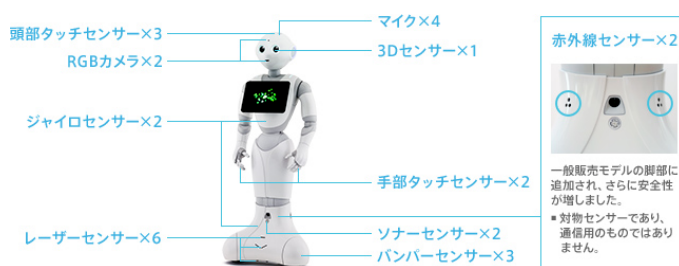


図 1: Pepper 外観

2. ロボットとのマルチモーダル対話

2.1 マルチモーダル情報の観測

ロボットは SoftBank 社の感情認識ヒューマノイドロボット Pepper^{*1} のデベロッパー先行モデルを使用する。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。図 1 に Pepper の外観と Pepper がもつセンサ機能を示す^{*2}。

2.2 POMDP

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。図 2 に POMDP のグラフィカルモデルを示す。

一般的に POMDP の観測状態は $\{S, A, T, O, Z, R, b_0\}$ 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

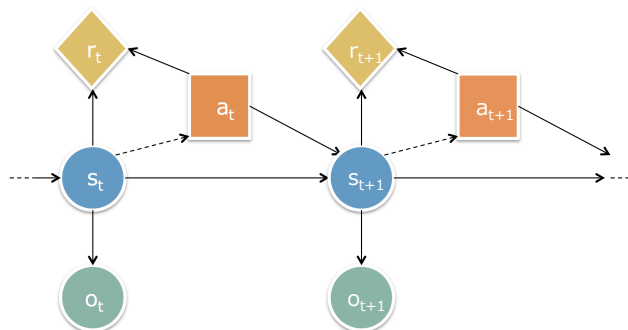


図 2: POMDP のグラフィカルモデル

測される観測値を表す。 Z は行動 a によって状態が 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 $b'(s')$ は式 (1) の漸化式で記述される。

$$b'(s') = k \cdot P(o'|s', a) \sum_s P(s'|s, a) b(s) \quad (1)$$

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

連絡先: 飯島采永, お茶の水女子大学, 〒112-8610 東京都文京区大塚 2-1-1, g1220503@is.ocha.ac.jp

^{*1} <http://www.softbank.jp/robot/>

^{*2} SoftBank 社のホームページ

(<http://www.softbank.jp/robot/consumer/products/spec/>) から引用した。

2.3 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 を考える。

- 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などの物理状態を示す。観測 o^p はレーザー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。

マルチモーダル対話例を表1に示す。

表 1: 対話例

話者	発話・行動	観測	$b(s)$
ユーザ Pepper	(遠くにいて) 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) 疲れた顔をしているね	o^e (悲しい)	

3. 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日分のデータに対する報告文の生成を行う。

3.1 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。

3.2 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

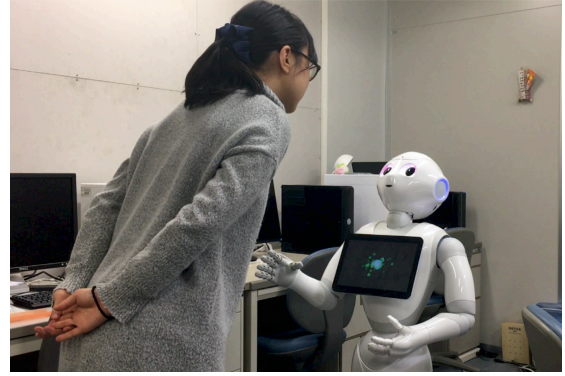


図 3: 実際の様子

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、観測の場面の变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考えて区切ることにする。

3.3 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」*3 と「飯」*4 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。

3.4 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、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

*3 <https://ja.wikipedia.org/wiki/天気>

*4 <https://ja.wikipedia.org/wiki/飯>

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	ijima	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	ijima	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	ijima	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	ijima	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	ijima	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	ijima	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	ijima	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	ijima	0.69	{ 0.02, 0.10, 0.51, 0.05, 0.32 }	-	-

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った場合に使用する。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, ijima, 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, ijima, 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, ijima, 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。」という文が生成できる。

4. まとめと今後の課題

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。

謝辞

本研究は立石科学振興財団の支援を受けました。ここに感謝致します。

参考文献

- [1] 岸本 康秀, 滝口 哲也, 有木 康雄 “階層的強化学習を適用した POMDP による音声対話制御”, 電子情報通信学会技術研究報告: 信学技報, pp. 121-126, 2010.
- [2] J. Pineau, G. Gordon, and S. Thrun “Point-based value iteration: An anytime algorithm for pomdps”, In Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1025-1032, 2003.
- [3] B. Bonet “An e-optimal grid-based algorithm for partially observable Markov decision processes”, In Proc. of ICML, pp. 51-58, 2002.
- [4] Jason D. Williams, Steve Young “Partially observable Markov decision processes for spoken dialog systems”, Computer Speech and Language, Volume 21, Issue 2, pp. 393-422, 2007.