

# 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

An Approach to realizing a Multimodal Dialogue System using a Robot

飯島采永<sup>\*1</sup> 小林一郎<sup>\*2</sup>  
Sae Iijima Ichiro Kobayashi

<sup>\*1</sup>お茶の水女子大学理学部情報科学科

Department of Information Sciences, Faculty of Science, Ochanomizu University

<sup>\*2</sup>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 recent years, with the spread of the household robots, it is considered necessary to enhance the communication capabilities of those robot. The objective of this study is to build a framework for a dialogue system dealing with multimodal information that a robot observes. We have applied partially observable Markov Decision Process to modeling multimodal interaction between a human and a robot. Through experiments we have confirmed that our proposed framework functions properly.

## 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、ロボットは複数のモダリティによる情報を観測し、それらの情報に基づいて適切な言動をすることが期待される。そこで、本研究では家庭用ロボット Pepper の身体性を利用したインタラクションの実現を目的とし、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 (POMDP) に基づくロボットとの対話処理に取り組む。

## 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<sup>\*1</sup> 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<sup>\*2</sup>。

### 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

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

<sup>\*1</sup> <http://www.softbank.jp/robot/>

<sup>\*2</sup> SoftBank 社のホームページ

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

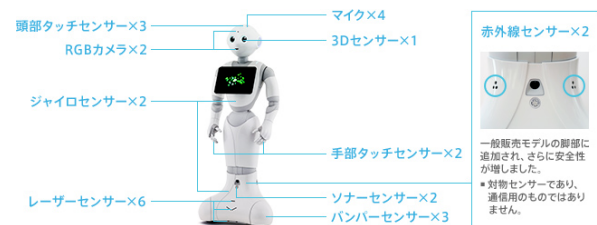


図 1: Pepper 外観

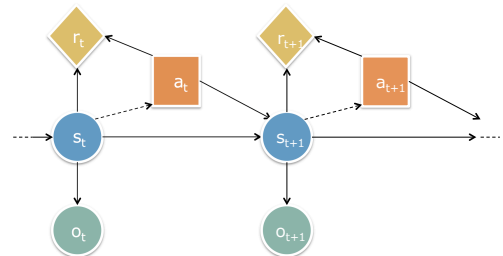


図 2: POMDP のグラフィカルモデル

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。

### 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:  $s^p(s^{p-dis}, s^{p-sense})$   
今回は Pepper のセンシングの機能から物理状態は距離と感覚の2つに分けることができる。それらはユーザがロボットからどれくらいの距離にいるのか ( $s^{p-dis}$ )、及び、ロボットに触っているかいないか ( $s^{p-sense}$ ) に相当する。観測  $o^{p-dis}$  はレーザーセンサやソナーセンサ、 $o^{p-sense}$  は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。

この状態に対応する観測をそれぞれ  $o^e, o^{p-dis}, o^{p-sense}, o^l$  とする。状態  $s \in S$  は  $s = (s^e, s^{p-sense}, s^l)$ , 観測  $o \in O$  を  $o = (o^e, o^{p-sense}, o^l)$  とする。観測  $o$  と状態  $s$  の対応関係を図3に示す。

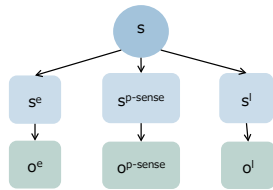


図3: 状態とマルチモーダル観測情報の関係

### 2.4 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[1]。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な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。

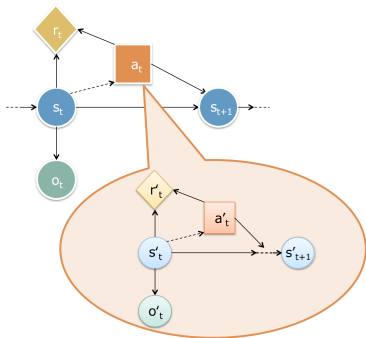


図4: POMDPの階層關係

### 2.5 Q学習による最適方策の獲得

ある状態  $s$  に対して行動  $a$  を選択する方針を方策  $\pi$  として定義する。 $\pi^*$  は  $s$  に対して最適な行動  $a^*$  を選択する最適

の方策である。本来、状態  $s$  は信念状態  $b(s)$  が確率値で表され、確定的ではないため、Point-Based Value Iteration[2] や Grid-Based Value Iteration[3] などを適用し、状態  $s$  を連続値として取り扱うことによる計算量の爆発を避けるが、本研究では、状態  $s$  は確定的であると仮定し、最適方策  $\pi^*$  は MDP を対象にした Q 学習で代用することにより求める。Q 学習の更新式は式(2)のようになっている。ここで、 $\alpha$  は学習率、 $\gamma$  は割引率を示す。

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r' + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (2)$$

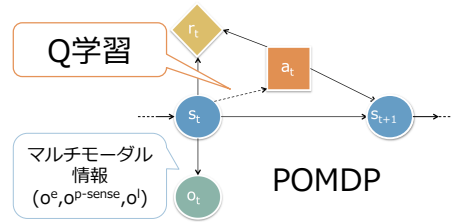


図5: POMDP と Q 学習の關係

## 3. マルチモーダル対話実験

マルチモーダル情報を使った Pepper との対話シナリオを POMDP の枠組みに沿って動かす。今回、Pepper が観測する物理状態のうち距離に関する状態が引き金となり対話の実行されるとみなし、距離状態の POMDP の中に対話の POMDP が階層的に含まれている対話シナリオを用いて実験を行った。POMDP には文献[4]を実装した、GitHub 上で公開されている Python のコード<sup>\*3</sup>を拡張して用いた。また、Pepper の制御には専用のソフトウェアとして公開されている PythonSDK<sup>\*4</sup>を利用した。

### 3.1 対話シナリオ

マルチモーダル対話シナリオを表1に示す。

表1: マルチモーダル対話シナリオ

話者	状態	観測情報	行動
ユーザ	(遠くにいる)	距離情報	
Pepper	近くにおいでよ		呼ぶ
ユーザ	(近くにくる)	距離情報	
Pepper	僕とお話しよう		話しかける
ユーザ	こんにちは	音声情報	
Pepper	こんにちは		挨拶する
ユーザ	(暗い顔)	画像情報	
Pepper	元気ないですね 僕が励まします		励ます
ユーザ	ありがとう (頭をなでる)	音声情報 センサ情報	
Pepper	照れるなあ		照れる
ユーザ	(いなくなる)	距離情報	
Pepper	(会話を終了する)		終わる

\*3 <https://github.com/mbforbes/py-pomdp>

\*4 <http://doc.aldebaran.com/1-14/dev/python/index.html>

### 3.2 実験設定

本実験では、状態の階層的関係をモデル化した POMDP を用いる。ユーザのいる位置を判定するタスクと、ユーザと対話を行うタスクに分解した。POMDP の設定は以下のように定義する。ただし、状態遷移確率、観測値出力確率、報酬、初期信念状態は手で設定をした。

- $S$  : ユーザ状態  
 $S^{p-dis} : \{ \text{いない, 遠い, 近い} \}$   
 $S : \{ \text{挨拶, 悲しい, 楽しい, 嬉しい, 嬉しくない} \}$
- $A$  : 行動  
 $A^{p-dis} : \{ \text{なにもしない, 終わる, 呼ぶ, 話しかける} \}$   
 $A : \{ \text{なにもしない, 挨拶する, 励ます, 楽しくなる, 照れる, へこむ} \}$   
 ここで、 $A^{p-dis}$  は  $S^{p-dis}$  に対応する行動である。
- $T$  : 状態遷移確率  $P(s'|s, a)$   
 表 2, 3 にある状態  $s$  から次の状態  $s'$  への遷移確率を示す。本実験では、どの行動に対しても同じ遷移確率を与えた。

表 2: 距離状態 : 状態遷移確率

$s^{p-dis} / s^{p-dis'}$	いない	遠い	近い
いない	0.2	0.15	0.15
遠い	0.2	0.15	0.15
近い	0.3	0.2	0.2

表 3: 対話状態 : 状態遷移確率

$s / s'$	挨拶	悲しい	楽しい	嬉しい	嬉しくない
挨拶	0.2	0.25	0.25	0.15	0.15
悲しい	0.2	0.15	0.15	0.25	0.25
楽しい	0.2	0.15	0.15	0.25	0.25
嬉しい	0.3	0.2	0.2	0.15	0.15
嬉しくない	0.3	0.2	0.2	0.15	0.15

- $O$  : 観測情報  
 $o^e$  : 画像情報による表情認識  
 $o^{p-dis}$  : 距離情報  
 $o^{p-sense}$  : タッチセンサ情報  
 $o^l$  : 音声情報
- $Z$  : 観測値出力確率  $P(o'|s', a), P(o^{p-dis}|s^{p-dis'}, a)$   
 Pepper はセンサを用いて観測値の識別率を算出することができるが、本研究では、音声情報を正しく観測する確率を 0.8, 画像情報とタッチセンサ情報、距離情報を正しく観測する確率を 0.7 とした。また、状態遷移確率と同様に、どの行動の後でも観測値出力確率は変わらないように設定した。
- $R$  : 報酬  $r(s, a)$   
 各行動後に逐次的に与える報酬。詳細を表 4, 5 に示す。
- $b_0$  : 初期信念状態  
 $b_0^{p-dis} = (\text{いない} : 0.2, \text{遠い} : 0.2, \text{近い} : 0.2)$   
 $b_0 = (\text{挨拶} : 0.3, \text{悲しい} : 0.2, \text{楽しい} : 0.2, \text{嬉しい} : 0.15, \text{嬉しくない} : 0.15)$

表 4:  $R^{p-dis}$  : 報酬

$s^{p-dis} / a^{p-dis}$	そのまま	終わる	呼ぶ	話しかける
いない	-1	5	-10	-10
遠い	-1	-10	5	-10
近い	-1	-10	-10	5

表 5:  $R$  : 報酬

$s / a$	なにもしない	挨拶する	励ます	楽しくなる	照れる	へこむ
挨拶	-1	5	-10	-10	-10	-10
悲しい	-1	-10	5	-10	-10	-10
楽しい	-1	-10	-10	5	-10	-10
嬉しい	-1	-10	-10	-10	5	-10
嬉しくない	-1	-10	-10	-10	-10	5

- $\pi^*$  : 最適方策  
 信念状態  $b(s)$  における最適行動  $a^*$  を示す最適方策  $\pi^*$  は Q 関数より式 (3) のようになる。

$$\pi^*(b(s)) = \underset{a}{\operatorname{argmax}} Q(b(s), a) \quad (3)$$

### 3.3 実験結果

実験結果を表 6, POMDP のグラフィカルモデルにのせたものを図 6 に示す。観測  $o$  より、状態  $s$  を推測する。信念状態  $b(s)$  から方策に従い、行動を選択している。また、行動価値の期待値を報酬  $r$  として出力した。ここで、Q 学習は  $\epsilon$ -greedy 法を使用し、学習率  $\alpha = 0.2$ , 割引率  $\gamma = 0.9$  を用いた。

### 3.4 考察

実験結果より、想定したシナリオを POMDP の枠組みに沿って動かすことができたことを確認した。また、対話システムにおける状態の表現は対象のロボットのセンシング機能にも依存してしまうことがわかった。本研究では物理的な距離を対話のきっかけととらえ全体の対話処理を構築したが、これは対象とするシステムによって様々なアプローチがあると考えられる。本実験では、状態を確定的であると仮定し最適方策を求めたが、今後、状態を連続的な状態として扱う必要がある。また、マルチモーダル情報の観測を逐次的に処理することしかできていないため、異なる複数のモダリティの情報が同時に与えられることで意思決定がされ、行動につなげるという枠組みも必要と考えられる。

## 4. まとめと今後の課題

本研究では、Pepper を対象にした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、距離情報を中心に階層的な状態構成に基づくインタラクションを行うことができた。今後の課題として、状態を連続的な状態として扱い、最適方策を求めることや、心理情報や言語情報についても階層的な状態構成を考えることが挙げられる。また、複数のモダリティが入力情報として同時に与えられた時の行動選択に対しても考察を行うつもりである。

### 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。

表 6: 実験結果

話者	状態	観測情報	信念状態 $b(s)$	行動	報酬										
ユーザ Pepper	(遠くにいる) (何もしない)	$o^{p-dis}$ [遠い]	<table border="1"><tr><td>0.182</td><td>0.545</td><td>0.273</td></tr><tr><td>いない</td><td>遠い</td><td>近い</td></tr></table>	0.182	0.545	0.273	いない	遠い	近い	なにもしない	-0.986				
0.182	0.545	0.273													
いない	遠い	近い													
ユーザ Pepper	(遠くにいる) 近くにいでよ	$o^{p-dis}$ [遠い]	<table border="1"><tr><td>0.082</td><td>0.735</td><td>0.184</td></tr><tr><td>いない</td><td>遠い</td><td>近い</td></tr></table>	0.082	0.735	0.184	いない	遠い	近い	呼ぶ	1.100				
0.082	0.735	0.184													
いない	遠い	近い													
ユーザ Pepper	(近くにくる) 僕とお話しよう	$o^{p-dis}$ [近い]	<table border="1"><tr><td>0.06</td><td>0.222</td><td>0.718</td></tr><tr><td>いない</td><td>遠い</td><td>近い</td></tr></table>	0.06	0.222	0.718	いない	遠い	近い	話しかける	0.819				
0.06	0.222	0.718													
いない	遠い	近い													
ユーザ Pepper	こんにちは こんにちは	$o^l$ [こんにちは]	<table border="1"><tr><td>0.8</td><td>0.05</td><td>0.05</td><td>0.05</td><td>0.05</td></tr><tr><td>挨拶</td><td>悲しい</td><td>楽しい</td><td>嬉しい</td><td>嬉しくない</td></tr></table>	0.8	0.05	0.05	0.05	0.05	挨拶	悲しい	楽しい	嬉しい	嬉しくない	挨拶する	2.541
0.8	0.05	0.05	0.05	0.05											
挨拶	悲しい	楽しい	嬉しい	嬉しくない											
ユーザ Pepper	(暗い顔) 元気ないですね 僕が励まします	$o^e$ [暗い顔]	<table border="1"><tr><td>0.727</td><td>0.156</td><td>0.035</td><td>0.035</td></tr><tr><td>挨拶</td><td>悲しい</td><td>楽しい</td><td>嬉しい</td><td>嬉しくない</td></tr></table>	0.727	0.156	0.035	0.035	挨拶	悲しい	楽しい	嬉しい	嬉しくない	励ます	1.539	
0.727	0.156	0.035	0.035												
挨拶	悲しい	楽しい	嬉しい	嬉しくない											
ユーザ Pepper	ありがとう (頭をなでる) 照れるなあ	$o^l$ [ありがとう] $o^{p-sense}$ [頭に触れる]	<table border="1"><tr><td>0.045</td><td>0.035</td><td>0.035</td><td>0.729</td><td>0.156</td></tr><tr><td>挨拶</td><td>悲しい</td><td>楽しい</td><td>嬉しい</td><td>嬉しくない</td></tr></table>	0.045	0.035	0.035	0.729	0.156	挨拶	悲しい	楽しい	嬉しい	嬉しくない	照れる	1.965
0.045	0.035	0.035	0.729	0.156											
挨拶	悲しい	楽しい	嬉しい	嬉しくない											
ユーザ Pepper	(いなくなる) (何もしない)	$o^{p-dis}$ [いない]	<table border="1"><tr><td>0.464</td><td>0.298</td><td>0.238</td></tr><tr><td>いない</td><td>遠い</td><td>近い</td></tr></table>	0.464	0.298	0.238	いない	遠い	近い	なにもしない	-0.989				
0.464	0.298	0.238													
いない	遠い	近い													
ユーザ Pepper	(いない) (会話を終了する)	$o^{p-dis}$ [いない]	<table border="1"><tr><td>0.796</td><td>0.146</td><td>0.058</td></tr><tr><td>いない</td><td>遠い</td><td>近い</td></tr></table>	0.796	0.146	0.058	いない	遠い	近い	終わる	1.964				
0.796	0.146	0.058													
いない	遠い	近い													

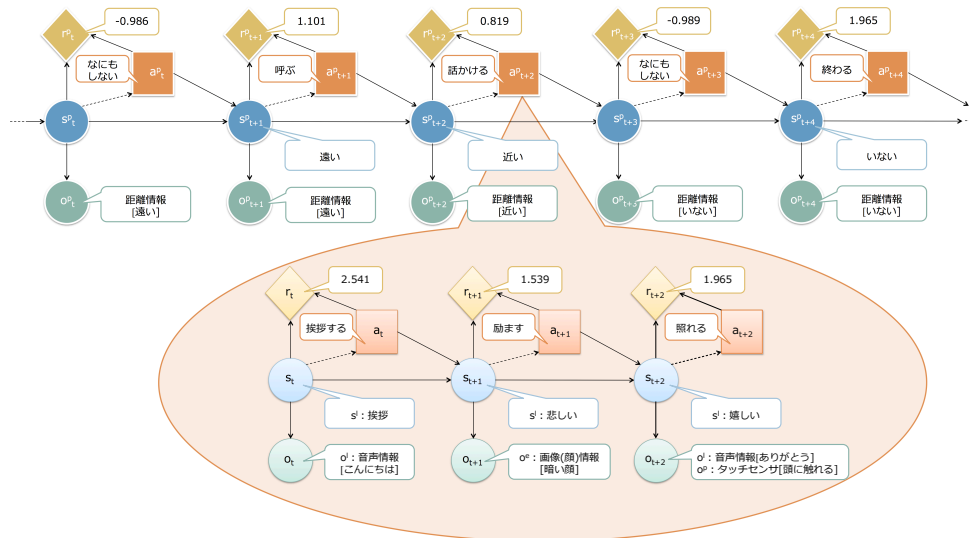


図 6: シナリオの POMDP

参考文献

- 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.
- 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.
- B. Bonet “An e-optimal grid-based algorithm for partially observable Markov decision processes”, In Proc. of ICML, pp. 51-58, 2002.
- 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.
- 木村元, Leslie Pack Kaelbling “部分観測マルコフ決定過程下での強化学習”, 人工知能学会誌, pp. 822-830, 1997.
- 盧迪, 久保 伸太郎, 深山 覚, 中沢 正幸, 西本 卓也, 嵯峨山 茂樹 “マルチモーダル入力と強化学習による擬人化エージェントの対話制御の検討”, The 24th Annual Conference of the Japanese Society for Artificial Intelligence, 2010.
- 南泰浩 “部分観測マルコフ決定過程を用いたインタラクション制御入門”, <http://www.lai.kyutech.ac.jp/sig-slud/SLUD63-minami-POMDP-tutorial.pdf>, 2011.
- 吉野幸一郎, 河原達也 “ユーザの焦点に適応的な雑談型音声情報案内システム”, 言語処理学会 第 20 回年次大会 発表論文集, pp. 761-764. 言語処理学会, 2014.