部分観測マルコフ決定過程に基づいたマルチモーダル対話への取り組み

飯島采永　(指導教員：小林一郎)

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 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．

2.2　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)[1] の枠組みを用いる．図 1 に POMDP のグラフィカルモデルを示す．

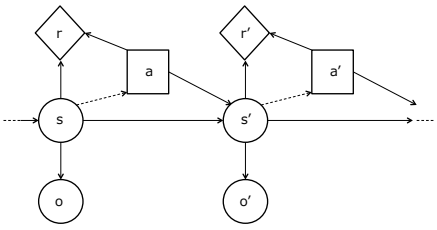

図 1: POMDP のグラフィカルモデル

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．

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 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い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，タッチセンサ等から取得する．
- 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．観測 $\boldsymbol{o} \in O$ を $\boldsymbol{o} = (o^e, o^p, o^l)$ とする．

2.4　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 π として定義する．π^* 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 $b(s)$ が確率値で表されるため確定的ではないが，本研究では状態 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ため，状態 s は確定的であると仮定し，最適方策 π^* は MDP を対象にした Q 学習 [2] で代用することにより求める．Q 学習の更新式は以下のようになっている．ここで，α は学習率，γ は割引率を示す．

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

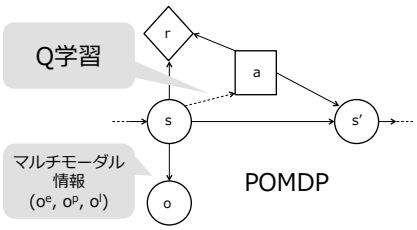

図 2: POMDP と Q 学習の関係

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．

3.1 実験設定

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[1] を利用した．POMDP の設定は以下のように定義する．ただし，今回の実験では状態遷移確率，観測値出力確率，報酬，初期信念状態は人手で設定をした．

- S：状態 { 挨拶, 悲しい, 楽しい, 嬉しい, 嬉しくない }
- A：行動 { なにもしない, 挨拶する, 励ます, 笑う, 照れる, へこむ }
- T：状態遷移確率 $P(s'|s,a)$
  表 1 に状態 s から s' への遷移確率を示す．ここでは，Q 学習によって得られた最適な行動に対する遷移確率を採用する．

表 1: 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 $\{\boldsymbol{o}_1, \boldsymbol{o}_2, ..., \boldsymbol{o}_t\}, \boldsymbol{o}_t = ({o_t}^e, {o_t}^p, {o_t}^l)$
  ${o}^e$ は画像情報による表情認識， ${o}^p$ はタッチセンサ情報，${o}^l$ は音声情報を表す．
- Z：観測値出力確率 $P(\boldsymbol{o}'|s',a)$
  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を正しく観測する確率を 0.7 とした．最適な行動の元での観測値出力確率を設定した．
- R：報酬 $r(s,a)$
  各行動後に逐次的に与える報酬．状態に対して正しい行動をした場合は+5，誤った行動をした場合には-10 の報酬を与えた．ただし，行動「なにもしない」を選択した場合は-1 の報酬を与えた．
- b_0：初期信念状態
  初期信念状態は，$b_0 =$ (挨拶：0.2, 悲しい：0.2, 楽しい：0.2, 嬉しい：0.2, 嬉しくない：0.2) とした．
- π^*：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 π^* は Q 関数より以下のようになる．

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

マルチモーダル対話シナリオを表 2 に示す．

表 2: マルチモーダル対話シナリオ

話者	発話内容	観測情報	行動
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

[1]http://doc.aldebaran.com/1-14/dev/python/index.html

3.2 実験結果

実験結果を表 4 に示す．観測 $\boldsymbol{o}$ より，状態 s を推測する．1 段目では, ユーザが「こんにちは」と発話したという観測に対し，$b(s) =$ (挨拶：0.8, 悲しい：0.05, 楽しい：0.05, 嬉しい：0.05, 嬉しくない：0.05) を得た．これにより確率の一番大きい「挨拶」という状態を推測した．信念状態 $b(s)$ から方策に従い,「挨拶する」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，学習率 $\alpha = 0.2$，割引率 $\gamma = 0.9$ を用いた．

表 3: 実験結果

話者	発話内容	実行結果
ユーザ Pepper	こんにちは こんにちは	$\boldsymbol{o} = (0, 0, o^l[\text{こんにちは}])$, $b(s) = (0.8, 0.05, 0.05, 0.05, 0.05)$, a:挨拶する，$r = 2.541$
ユーザ Pepper	(暗い顔) 元気ないですね 僕が励まします	$\boldsymbol{o} = (0, o^e[\text{暗い顔}], 0)$, $b(s) = (0.046, 0.727, 0.156, 0.035, 0.035)$, a:励ます，$r = 1.539$
ユーザ Pepper	ありがとう (頭をなでる) 照れるなあ	$\boldsymbol{o}$= $(0, o^p[\text{頭に触れる}], o^l[\text{ありがとう}])$, $b(s) = (0.045, 0.035, 0.035, 0.729, 0.156)$, a:照れる，$r = 1.965$

3.3 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．
しかし，今回の実験では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．例えば，ユーザが「ありがとう」と発話し,「頭をなでる」という行動を観測するのに，まず音声情報を観測し，その後タッチセンサ情報を観測している．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が必要と考えられる．

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．今後の課題として，今回は人手で設定した状態遷移確率，観測値出力確率を自動で獲得す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．

参考文献

[1] Kaelbling, L.P., Littman, M.L., Cassandra, A.R. “Planning and acting in partially observable stochastic domains”. Artificial Intelligence Journal 101,pp. 99-134, 1998.

[2] Richard Sutton, Andrew Barto,“Reinforcement Learning: An Introduction”, MIT Press, 1998.

[3] 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.

[4] 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

[5] 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.