

ソマティック・マーカー仮説に基づく行動選択

Action Selection based on Somatic Marker Hypothesis

日永田智絵^{*1}

Chie Hieida

堀井隆斗^{*1}

Takato Horii

長井隆行^{*1*2}

Takayuki Nagai

^{*1}電気通信大学

The University of Electro-Communications

^{*2}大阪大学

Osaka University

Emotions are very important for human intelligence; however the mechanism of emotions is not yet fully clarified. The important aspect of embodiment in emotion has been claimed by Damasio's somatic marker hypothesis, which proposed that emotions evaluate external stimuli efficiently through the body. As a first step toward understanding the mechanism of emotion, we try to verify the somatic-marker hypothesis using a computer simulation. Specifically, we introduce a module that learns actions using body signals, and verify whether the agent can learn to obtain higher reward using the body signals. As the result, the simulation reveals that the model with body signals can select actions with higher reward compared to models without signals from the body.

1. はじめに

感情は人において重要な要素だといわれる。近年の深層学習の発展に伴い、人工物は人を凌ぐ計算能力を有するようになるといわれるものの、感情は人工物には有することのできないものであるという考え方が多い。しかしながら、意識が脳の計算の結果、生じているものだとするならば、感情も計算機上で表現できる可能性があるのではないだろうか。

感情に関してはその重要な役割から、過去にも多く感情に関する研究が行われてきた。その中の重要な仮説として、ソマティック・マーカー仮説というものが存在する。ソマティック・マーカー仮説とは、神経科学者の Damasio が、情動を身体シグナル、その認知を感情とし、外部の刺激によって引き起こされる身体シグナルが意思決定を効率化しているとした仮説である [Damasio 96]。

本研究では、感情メカニズム解明への足掛かりとして、深層学習を用い、ソマティック・マーカー仮説を中心とした感情の統合的なシステムを構築し、ソマティック・マーカー仮説の検証を行うこととした。具体的な検証方法は、意思決定の効率化を強化学習における報酬の高い行動選択が行えるようになることと解釈し、深層学習を用い、行動方策を学習するネットワークを構築し、身体シグナルがあることによって、報酬の高い行動の選択が行われるかを検証した。

本研究の意義は、感情の統合的なシステムを構築し、その振る舞いを考察することによって感情メカニズム解明を目指すアプローチ（構成的アプローチ）を行う点にある。例えば生理学的研究では、ジェームズ・ランゲ説を初めとした感情の理論が提案されており [James 84]、神経科学では、各脳部位の感情的な働きを調査している [Koelsch 15]。これらの研究では、一部分のモデルや各部位に絞った研究となっており、感情の統合的なメカニズムについてはまた完全には明らかになっていない。各部位に焦点を当てる方法は、その部位を知るために重要であるが、統合的なシステムの中で初めて現れる機能もあると考えられる。そのため、感情の統合的なシステムを構築し、その振る舞いを考察することによって感情メカニズム解明を目指すアプローチも促進されるべきである。

2. 実験プロトコル

仮説検証のため、まずは人の身体との対応を考慮したシンプルなタスクを設計した。本研究では、射的タスクを題材にした（図 1）。タスクには水鉄砲、バケツ、風船が存在し、それぞれ身体反応（筋肉の動きや発汗、心拍など）、体力、外的刺激（食料）を表し、エネルギーの流れは水の移動で表現される。バケツと水鉄砲に入った水は時間が経つと蒸発してしまう。3 回に 1 回、風船が飛んでくるという状況で、行動として水鉄砲を打つか、バケツの蓋を抑えるかを選択することができる。風船を水鉄砲で打つことが出来れば、風船から水を獲得することができ、バケツの水が増える。バケツの蓋を抑えると蒸発量を減らすことができる。水鉄砲に水を増やす際にはバケツの水を減らすこととなる。このような状況下で自身の持つ水の量を一定に保つように行動を学習する。この設定はホメオスタシスの考え方からである。その際に身体反応と体力の情報即ち身体シグナルが行動決定を効率化するかということを検証することになる。身体シグナルを出力するモジュールを本研究では、Somatic Marker Module (SMM) と呼ぶ。タスク内の詳細値は以下に示す。

- 初期値は水鉄砲が 0、バケツが 80。
- 値の範囲は水鉄砲が 0-10、バケツが 0-100。
- 水鉄砲とバケツの水の総量を 80 に保つことが目標。
- 打つ行動を選択すると、水鉄砲-10。
- 但し水鉄砲が 10 に満たない場合は打つことができない。
- 風船を打つことができると、バケツ+30。
- 画像が変わるごとに水が蒸発。水鉄砲-2、バケツ-2。
- 但し蓋を抑える行動を選択するとバケツの水の蒸発が半減。

3. 深層学習による実装

実装したネットワークを図 2 に示す。本研究では、強化学習での行動選択を想定し、Deep Deterministic Policy Gradient (DDPG) [Lillicrap 15] を用いて、行動モジュールを実装した。また、時系列情報を扱えるようにするため、convolutional LSTM (long short-term memory) [Xingjian 15] を用いて、次の状況を予測するネットワークを追加した。本タスクにおいてはオーバーサペックであるとも考えられるが、さらに複雑な

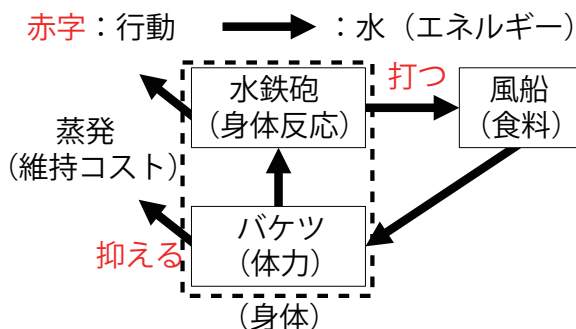


図 1: タスク設計概要

タスクを行うことを視野にいれ、本手法を選択した。身体シグナルを出力する Somatic Marker Module に関してはルールベースで作成した。

3.1 Somatic Marker Module (SMM)

SMM の具体的な役割は外的刺激と行動によって変化した現在の身体シグナルを受け、身体反応を引き起こし、行動モジュールに身体シグナルを入力することである。外的刺激を受け身体反応を引き起こす部分は、風船があるときには、水鉄砲に 10 用意することとした。水鉄砲に水を用意するにはバケツから足りない分の水を移動させることとなる。その操作をした後、次のモジュールに水鉄砲とバケツの値を入力する。

3.2 Convolutional LSTM

Convolutional LSTM は、Xingjian らによって提案された画像の特徴を捉える Convolutional Neural Network と長期時系列情報を扱うことができる LSTM とを組み合わせた方法である [Xingjian 15]。具体的には、LSTM の重みの乗算が畳み込みであり、構成要素はメモリセル C_t 、入力ゲート i_t 、忘却ゲート f_t 、出力ゲート o_t である。本稿では 2 層の Convolutional LSTM を用いており、フィルタは $5 \times 5 \times 5$ であり、誤差は平均二乗誤差によって計算される。学習率は Adam ($\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$) である。入力 は 外部刺激画像と身体シグナル ($65 \times 65 \times 5$ 次元) で、出力は予測画像と予測身体シグナル ($65 \times 65 \times 5$ 次元) である。身体シグナルは値をグレースケール画像化して入力している。

3.3 Deep deterministic policy gradient

Deep deterministic policy gradient (DDPG) は、Lillicrap らが提案した深層学習を用いた強化学習法であり、高次元の連続動作空間でポリシーを学習できる深層関数近似を用いたモデルフリー、オフポリシーの actor-critic アルゴリズムである [Lillicrap 15]。入力 は 外部刺激画像と LSTM による予測画像、身体シグナルと身体シグナルの初期値 ($65 \times 65 \times 12$ 次元) で、出力は行動 (1 次元) である。中間層は 300 次元の入出力で構成されている。報酬はホメオスタシスという考えを用い、下記のように算出する。

$$R(t) = C - \|\mathbf{m}(t)\| - \|\mathbf{a}'(t)\|, \quad (1)$$

$R(t)$ は時間 t における報酬値、 $\mathbf{a}(t)$ は時間 t における SMM の出力、すなわち身体シグナルを表す。それに対し、 $\mathbf{a}'(t)$ は $\mathbf{a}(t)$ に行動後の変化 (例えば風船から水が得られたなど) を加味したものとなる。 $\mathbf{m}(t)$ は時間 t での保つべき基準の値となって

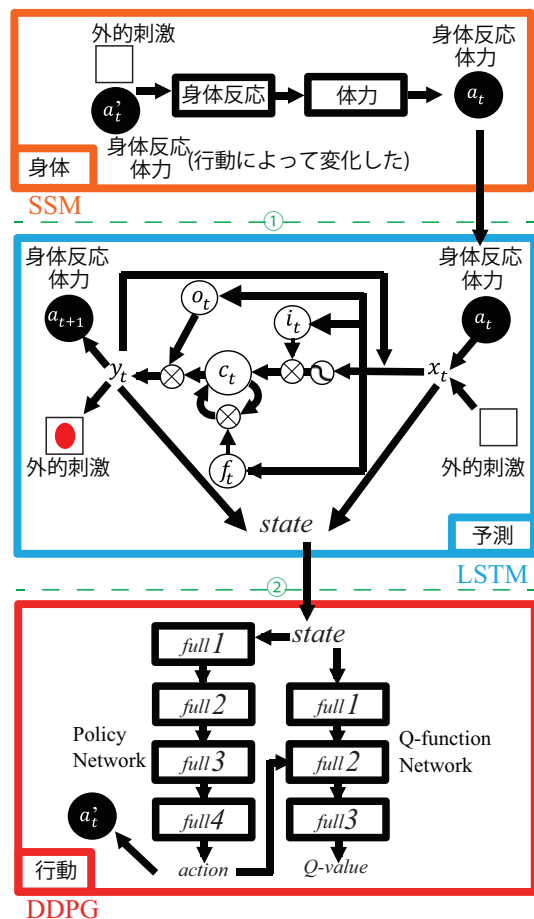


図 2: 実装したネットワーク

おり、本研究では初期値の値 $\mathbf{a}(0)$ とする。 C は報酬値の補正をする定数であり、 $R(t)$ は 0-100 の値をとる。但し、 $\|\mathbf{a}'(t)\|$ が 0 のとき、 $R(t)$ も 0 とし、 $\mathbf{a}'(t)$ の値を初期値に戻す。

学習率は Adam (actor network: $\alpha = 10^{-4}, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$, critic network: $\alpha = 10^{-3}, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$) となっており、探索手法は Ornstein-Uhlenbeck process、ミニバッチは 200、リプレイバッファのサイズは 500 で、新しいデータが入ると古いデータは破棄される。また、本研究では batch normalization を使用した。

4. 結果・考察

SMM のある条件を SMM 条件とし、それに対して SMM と LSTM が無い条件を DDPG 条件 (図 2 の②の破線でネットワークが切れている条件)、SMM が無い条件を DDPG-LSTM 条件 (図 2 の①の破線でネットワークが切れている条件) と呼ぶ。欠損した情報は 0 で入力している。学習時の報酬の様子を図 3 に示す。学習はそれぞれ 20000 epochs 行った。グラフは見やすくするために 50 epochs ごとの平均値をプロットしている。報酬のグラフより、全ての条件で報酬は高くなっているが、身体シグナルありの場合がより高い報酬が得られていることがわかる。総時間の報酬の平均値をとると、SMM 条件が 88.9、DDPG 条件が 83.1、DDPG-LSTM 条件が 83.2 となっており、身体シグナルありが最も大きい。しかし、この結果には学習時の行動のノイズも含まれているため、学習済みモデル

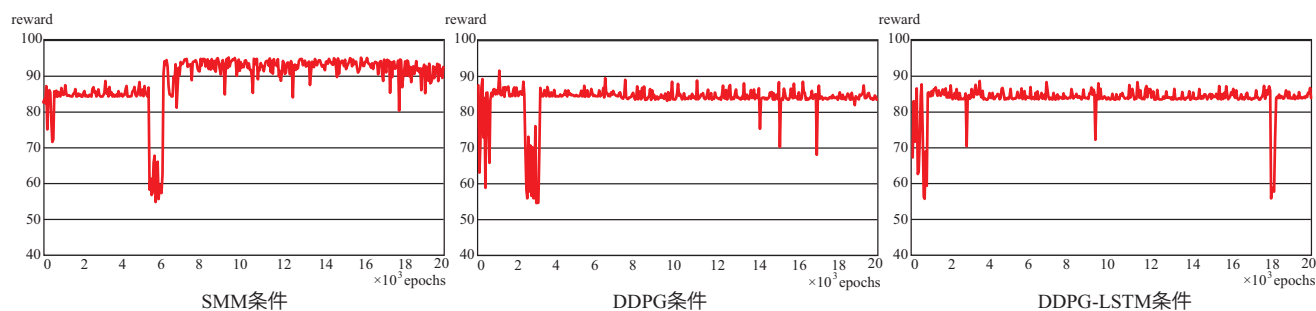


図 3: 学習時の報酬の推移

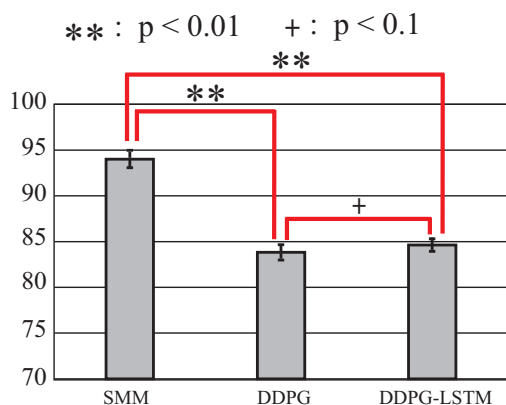


図 4: 学習済みモデルでの報酬平均の比較

に対し、それぞれ 20episode 動かした結果を平均したものと比較した。

結果を図 4 に示す。平均値は SMM 条件が 94.1, DDPG 条件が 84, LSTM 条件が 84.7 となった。ボンフェローニ法で解析を行った結果、SMM 条件と DDPG 条件、DDPG-LSTM 条件の間で有意差が見られた ($p = 2.94 \times 10^{-8}$, 2.16×10^{-7} , ボンフェローニ法より $p < 0.01/3$)。DDPG 条件と DDPG-LSTM 条件の間は有意傾向であった ($p = 0.027$, ボンフェローニ法より $p < 0.1/3$)。よって、身体シグナルありの場合が報酬を高くする行動を選択できることがわかる。また、DDPG と DDPG-LSTM の間は有意傾向であるので、予測モジュールが報酬増加に寄与したと考えられる。

LSTM の Loss のグラフは図 5 の通りである。グラフは見やすくするために 100epochs ごとの平均値をプロットしている。初期状態に比べ誤差が減少しており、学習が行われていることがわかる。最終的な誤差は 3.24 と 1.32 となり、DDPG-LSTM 条件の方が小さくなる。これは、身体シグナルの予測を行っていないためであると考えられる。

次に LSTM による予測の一例を図 6 に示す。画像の予測では、綺麗に画像の再現ができていないわけではないものの、風船が空の色になっていたり、空が風船の色になっているなど、多少の予測ができていたことがわかる。水量の予測では、非常に近い値ではないものの、入力された情報に対し値が増加するか減少するかというレベルでの予測はできている。

また、身体反応のパターンを変更し、学習した結果の行動と比較した。a 群は風船がきたときのみ水鉄砲に 10 用意する群で、b 群は画像に関わらず常に水鉄砲に 10 用意するパター

ンである。その結果 SMM 条件と DDPG-LSTM 条件では行動に違いが出たが、DDPG 条件では行動に違いが出なかった。これは、DDPG 条件では風船がどのタイミングで来るか予測できず、来た際に行動するというしかできないため、a 群と b 群に違いが出なかったと考えられる。予測モジュールがあると、風船が来ないパターンの状況を細分化できるため、行動にバリエーションが生まれる。しかしながら、詳細な身体状況はわからないため、周期的な行動に落ち着くこととなる。身体シグナルがあると、その時の身体状況に合わせて行動を選択することが可能となる。

また、SMM 条件の b 群の報酬平均は 93.7 で、DDPG 条件の報酬平均は 86, DDPG-LSTM 条件の報酬平均は 93.1 であった。SMM 条件と DDPG-LSTM 条件の報酬があまり変わらないのは、b 群のシナリオが身体反応を一定にしたために、身体シグナルなしでも十分に行動できたためだと思われる。しかしながら、打つ行動を選択し行動できている回数を比較すると、SMM 条件は 6 回で DDPG-LSTM 条件は 11 回となる。つまり、身体シグナルがあることにより、使用する水の量をより少なくより高い報酬を得ることができている。SMM 条件内で比較すると、a 群は打つことができているのが 2 回、b 群が 6 回となる。つまり、a 群の方が使用する水の量を少なくでき、身体反応のパターンが行動に対して制約をかけ、行動を効率化させていることがわかる。即ち、行動選択において、身体反応は重要であり、この部分をどのように設定するのかが問題となる。人においては、生得的な身体反応があると考えられており、その要素がエクマンの基本感情 [Ekman 71] の考えに繋がっていると考えられる。ただし、生得的な部分のみでは説明がつかない要素もあり、後天的に身体反応を学習するモジュールも必要であろう。

5. まとめ

本研究では、感情メカニズム解明への足掛かりとして、ソマティック・マーカー仮説の検証を行った。手法として、強化学習での行動選択を想定し、Deep Deterministic Policy Gradient (DDPG) を用いて行動を強化学習するモジュールを作成し、convolutional LSTM (long short-term memory) で作成した予測モジュール及び身体シグナルを出力する Somatic Marker Module (SMM) の有無によって、行動選択がどのように変容するかを検証した。学習済みモデルで行動した際の報酬の平均を比較した結果、身体シグナルありの SMM 条件が最も報酬が高い行動を選択できていることがわかった。また、身体反応のパターンの変化による検証では、身体反応のパターンが行動に対して制約をかけ、行動を効率化させていることがわかった。

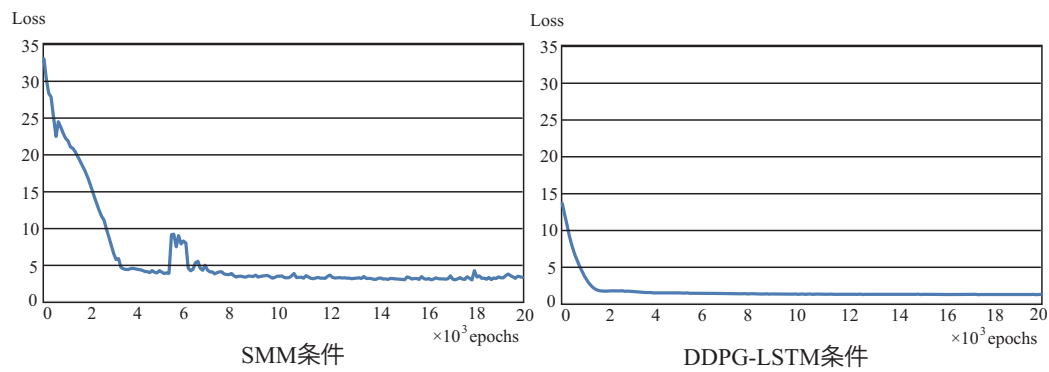


図 5: 学習時の LSTM の Loss の推移

表 1: 画像に対する反応を変更した際の行動例：画像 0 は風船なし，画像 1 は風船あり，行動 0 は蓋を抑える，行動 1 は打つ．

画像			0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
行動	SMM 条件	a 群	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	1	0	0	0
		b 群	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0
	DDPG 条件	a 群	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
		b 群	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
	DDPG-LSTM 条件	a 群	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	0	1
		b 群	1	1	1	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1	0

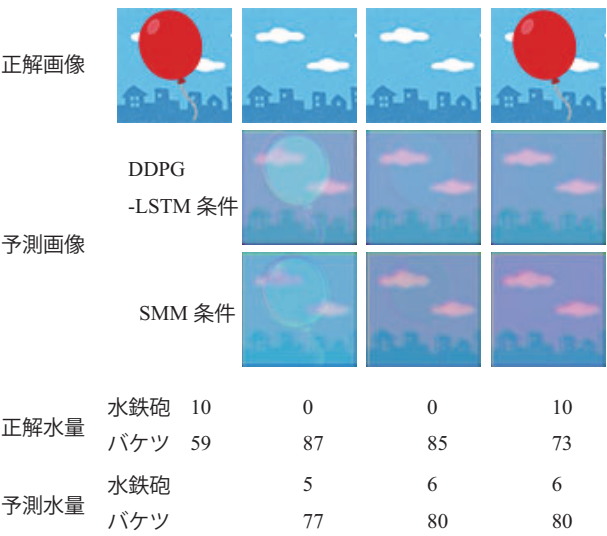


図 6: LSTM による予測の一例

これらは，身体シグナルが意思決定を効率化しているというソマティック・マーカー仮説の主張に沿った結果となっている．
今後は SMM-LSTM-DDPG の全体のネットワークに対し，より複雑なタスクを行い，挙動を確かめるとともに，社会的感情など，複雑な感情がどのように表現されるのかを検証する．

謝辞

本研究は JSPS 科研費 JP16J04930, JST CREST (JP-MJCR15E3)，新学術領域「認知的インタラクションデザイン学」(26118001)の助成を受けたものです．

参考文献

[Damasio 96] Damasio, A. R., Everitt, B. J., and Bishop, D.: The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion], *Philosophical Transactions of the Royal Society B, Biological Sciences*, Vol. 351, No. 1346, pp. 1413–1420 (1996)

[Ekman 71] Ekman, P. and Wallace, F. V.: Constants across cultures in the face and emotion, *Journal of personality and social psychology*, Vol. 17, No. 2, pp. 124–129 (1971)

[James 84] James, W.: WHAT IS AN EMOTION ?, *Mind*, Vol. os-IX, No. 34, pp. 188–205 (1884)

[Koelsch 15] Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., Scheve, von C., and Gebauer, G.: The quartet theory of human emotions: An integrative and neurofunctional model, *Physics of Life Reviews*, Vol. 13, pp. 1–27 (2015)

[Lillicrap 15] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D.: Continuous control with deep reinforcement learning, *arXiv preprint arXiv:1509.02971* (2015)

[Xingjian 15] Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in *Advances in neural information processing systems*, pp. 802–810 (2015)