The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

感情と記号創発ロボティクス

Emotions and Symbol Emergence in Robotics

長井 隆行 *1*2
Takayuki Nagai

*1大阪大学 Osaka University
*2電気通信大学 The University of Electro-Communications

Recent emotion studies reveal that interoception is a key concept to understand the mechanism behind human emotions. This idea can explain emergence of emotions, which can be seen as the differentiation mechanism of emotions. This paper tries to integrate the idea of emergence of emotions with the framework of symbol emergence in robotics. A preliminary study on grounding of affective words are demonstrated in this paper.

1. はじめに

感情とは何か？この問いは，古くから我々人間を悩ませ続けてきた．古くは哲学から，近年では心理学や神経科学をはじめロボティクスなど様々な領域で，この問いに対する答えの探求が続けられてきた．その探求は形を変えながら続けられているが，いまだにこの問いに対する明確な答えは存在しない．そうした中で近年の感情研究における一つの重要な示唆は，内受容感覚と情動の関係である [大平 17, 乾 18, Feldman17]．内受容感覚とは，内臓など身体内部の信号の処理の結果生じる知覚であるとされる．特に，内受容感覚の予測符号化は，予測に基づく脳の基本的な動作原理によって情動や感情を説明できる可能性を秘めている．ここで重要なのは身体であり，身体からの信号を自己組織化し自身の身体に関する予測性を高めることがポイントである．そうした情報の構造化やその構造に基づく意思決定・行動といった一連の人と環境の相互作用プロセスの中で感情が創発するものと考えられる．

本稿では，情動と感情を次のように定義する．情動とは，刺激に対する身体の反射的な反応を知覚した内受容感覚であり，体の状態を把握するための非常に基本的なものである．感情は，情動がより複雑に発展し，社会的文脈の中で発現したものであるが，内的な知覚であることに変わりはない．

ここで工学分野に目を向けてみると，人間のようなロボットやエージェントを作るという意味でも，感情は重要な要素として扱われてきた．特に感情を表出するという文脈で，多くの研究がなされている [Breazeal03]．こうした研究で問題となるのは，人がエージェントの表出をどのように捉えるかという人の心理的側面であり，このこと自体が必ずしも感情とは何かという問いに直接的に答えたり，感情を持ったエージェントを作ることを狙ったものではない．情動や感情の創発や分化という側面を捉えようとした研究は，それほど多くはない [浅田 15, 堀井 13]．

一方で知能ロボティクスでは，言語を含む知の創発的側面を扱う記号創発ロボティクスと呼ばれる領域が立ち上がっている．記号創発ロボティクスでは，ロボットやエージェントが環境や他者との相互作用を通して概念や言語をボトムアップに獲得すると考える．そうした複数の個体の知能が，ある種社会的な全体として記号システムを形成し，その記号システムがトップダウンに個体の学習を制約するという，ミクロ・マクロループとして我々の社会をとらえる．こうした社会を構成することのできるエージェントの計算モデルを構築し，物理的なロボットとして実現することが大きな目的である [Taniguchi16]．これまで記号創発ロボティクスでは，センサモータ情報の随伴性に基づいて，言語を獲得するタスクが主に検討されてきた [Nakamura11, Attamimi14]．もし情動や感情が内受容感覚を基盤とした創発現象だとすれば，これらも同じ枠組みに入れられるはずである．つまり記号創発ロボティクスで議論される計算モデルの枠組みに情動や感情を取り入れ，知能全体として議論することで，感情を持ったロボットを実現するための基礎的な検討ができる可能性がある．本稿では，こうした視点で情動/感情がどのようにモデル化され得るかについて議論することを目的とする．

連絡先: 長井隆行，大阪大学，大阪府豊中市待兼山町 1-3，06-6850-6365，nagai@sys.es.osaka-u.ac.jp

2. 感情モデル

ここではまず，感情の計算モデルを考える前段として，感情を含む知能の全体をどのように捉えるかについて考える．既に述べたように，感情を内受容感覚を基盤とした創発的な現象であるとするならば，以下の 3 点がポイントである．

- 身体の必要性
- 知能の枠組みの一つ（不可分性）
- 社会的側面の重要性

これらのことは，自動運転車が複数走っている様子に例えるとわかり易い．自動運転車には，自動車を制御するためのコンピュータが搭載されている．コンピュータには各種のセンサが接続されており，外界の情報が入力される．この情報は，自己位置の推定や障害物検知，経路の計画などに利用される．しかし，自動運転車にとって重要なセンサ情報はそれだけではない．走るための基盤として最も重要なのは，車自身の情報である．それは，現在のエンジンや燃料，タイヤの状態などである．燃料の情報は完全に正確に把握できるわけではなく，センシングに伴うノイズの影響が少なからずある．エンジンやタイヤの情報なども同様である．車は，こうした情報に基づいて走行に支障がないかどうかを瞬時に判断する必要がある．また，残りの可能な走行距離といった情報はこうしたセンシングに基づく予測であり，これから先どのような道をどれくらいのスピードで走るのかという不確定な要素を多分に含んでいる．この予測の精度を上げるためには，車がこれまでに取得した様々

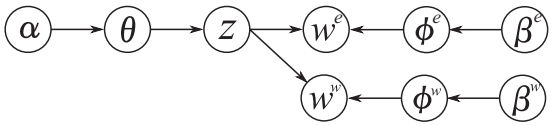

図 1: MLDA のグラフィカルモデル

なデータを構造化し，それらを総合的に参照する必要がある．このように，車においても自身の身体の情報は重要であり，それを知覚したものが情動であると捉えることで，主観的な意識の問題はあるものの，人間における情動の存在意義やメカニズムが考えやすくなる．

自動運転車の例でさらに重要なのは，他の自動運転車の存在である．外界の情報としての単なる障害物は，物理的な法則に従って存在するために，比較的その挙動は予測しやすい．しかし他の自動車は，そうした単なる物体とは異なる挙動を示す．むしろ他の個体は，自身を駆動するモデルを当てはめて予測した方が予測精度を高めることができるはずである．これは，自他分離によって，自身の予測モデルを他者の予測に利用する他者モデルの考え方であると言える．そしてこれらのことはすべて内的な知覚とも関係しているため，情動は外界に存在する物体や他者との関係として学習されることになる．こうして情動は他者との関係性や時間的な軸で複雑化することになり，これが感情であると考える．そこには，他者への羨望や将来への不安といった，比較的複雑ないわゆる高次感情が発現すると考えられる．

3. マルチモーダルカテゴリゼーションと情動

上述のような全体像を実現する前段として，情動を創り出すための非常に単純な例を示したい．以下では，内受容感覚が 2 次元のベクトルとして内部表現されていると仮定する．そしてこの 2 次元のベクトルはそれぞれ，感情価と覚醒度に対応していると仮定する．

3.1 感情的語彙の接地

筆者らのグループでは，確率的生成モデルの枠組みでマルチモーダル情報を構造化することで，ロボットが経験から概念を獲得し，さらに語彙との結びつけを相互に学習することで，語意を獲得できる可能性があることを示してきた [Nakamura11]. この仕組みをマルチモーダルカテゴリゼーションと呼ぶが，これに内受容感覚，つまりは身体の情報を入力することで，感情的語彙を獲得することを考える．マルチモーダルカテゴリゼーションは，階層ベイズや深層学習によって実装することができるが，ここでは階層ベイズに基づいたマルチモーダル LDA (MLDA: multimodal latent Dirichlet allocation) を用いた例を示す．

ここでの情動概念形成の単純化したモデルは，図 1 のグラフィカルモデルに相当する．このモデルにおいて，各観測情報 w^* は，それぞれハイパーパラメータ β^* によって決まるディリクレ事前分布に従うパラメータ ϕ^* の多項分布によって発生する．また z はカテゴリを示し，ハイパーパラメータ α によって決まるディリクレ事前分布に従うパラメータ θ の多項分布により発生する．

本稿におけるカテゴリ分類は，実際に取得した内的知覚情報と単語情報に基づき，モデルのパラメータ θ および ϕ を推定することに相当し，パラメータ推定には Gibbs Sampling を適用する．各知覚情報 m (情動の場合は感情価と覚醒度) のそれぞれの集合を $\boldsymbol{w}^m$，m 番目の情報の次元数を W^m とすると，対象における m 番目の情報の i 番目に割り当てられるカテゴリ z_{mi} は，

$$P\left(z_{mi}=k|\boldsymbol{z}^{-mi},\boldsymbol{w}^m,\alpha,\beta^m\right)\propto (N_k^{-mi}+\alpha)\cdot\frac{N_{mw^mk}^{-mi}+\beta^m}{N_{mk}^{-mi}+W^m\beta^m} \quad (1)$$

に従ってサンプリングされる．ここで N_{mw^mk} は，対象における m の情報 $\boldsymbol{w}^m$ についてカテゴリ k が割り当てられた回数を示す．式 (1) に従い，繰り返しサンプリングを行うことで結果がある値 $\hat{N}_*$ へと収束する．収束結果より，カテゴリの総数を K とする時，最終的なパラメータの推定値 $\hat{\phi}^m_{w^mk}$，$\hat{\theta}_k$ は以下のようになる．

$$\hat{\phi}^m_{w^mk}=\frac{\hat{N}_{mw^mk}+\beta^m}{\hat{N}_{mk}+W^m\beta^m},\quad \hat{\theta}_k=\frac{\hat{N}_k+\alpha}{\sum_k\hat{N}_k+K\alpha} \quad (2)$$

3.2 未知のシーンに対する語意の予測

学習したモデルを利用することで，未観測のシーンに対する情報の予測が可能である．本稿では，未知シーンを観測したことによって生成される情動信号 (感情価と覚醒度) に基づき単語を予測することを考える．観測した情動信号 $\boldsymbol{w}^e_{obs}$ から予測されるカテゴリ z，および単語 w^w は，それぞれ

$$z=\underset{z}{\mathrm{argmax}}\int P(z|\theta)P(\theta|\boldsymbol{w}^e_{obs})d\theta \quad (3)$$

$$P(\boldsymbol{w}^w|\boldsymbol{w}^e_{obs})=\int\sum_z P(w^w|z)P(z|\theta)P(\theta|\boldsymbol{w}^e_{obs})d\theta \quad (4)$$

によって求められる．これは，学習時に推定した情動信号を生成する多項分布のパラメータ ϕ^e を固定し，パラメータ θ を Gibbs Sampling により推定することで計算することが可能である．

3.3 語彙獲得の例

実際に IAPS を用いて行った実験の結果を示す．まず，IAPS[JLang05] を用いて畳み込みニューラルネットワーク (CNN) を用いて，入力画像に対して情動価と覚醒度を推定するネットワークを構築した．また，感情価と覚醒度によって簡単な表情を表出するシンプルな顔エージェントを作成した．次に，IAPS の画像と，CNN によって推定された値を用いて駆動したエージェントの顔を被験者に提示し，対応する複数の文書を自由に記述してもらった．図 2 に，データの例を示す．こうして得たデータを，前述の MLDA に入力することで，書く語彙が情動空間とどのように結びつき，モデルの内部でどのような潜在空間ができるかを調べた．

分類結果と確率的に結びついた語彙を，図 3 に示す．これより，感情価 (valence) が小さい左側のカテゴリにはネガティブな感情語が結びついており，逆に右側のカテゴリにはポジティブな語が結びついていることが分かる．学習データの文章は，感情語だけでなく，女性や子どもなどの一般名詞が含まれているため，例えば (e) のカテゴリには，女性という単語が結びついている．これは，感情価が高くかつ覚醒度の高い画像データの多くが，女性を含んでいたためである．こうしたカテゴリ分類がなされてしまうと，感情価が高くかつ覚醒度の高い状況で「女性」という単語が高い確率で予測されてしまい問題となる．画像中に何が写っているかという文脈と情動は分離してモデル化されるべきであるが，これを実現するためには，物体概念など多様な概念を別々に学習し，それらを階層的に結合する多層 MLDA[Attamimi14] を用いる必要がある．

図 2: 学習データの例

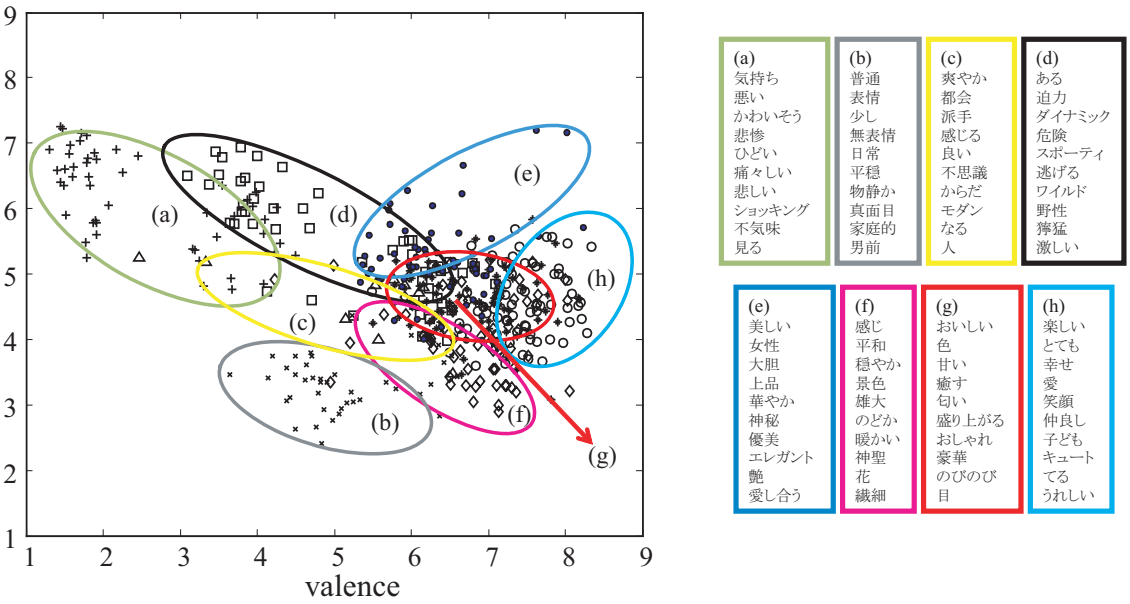

図 3: 分類 (学習) の結果

3.4 未知のシーンに対する単語予測

前述の IAPS データを用いて学習したモデルに対して，学習に用いていない画像を入力し，単語予測の実験を行った．予測結果の例を，図 4 に示す．各図は入力画像と，その際のロボットの表情，および単語予測結果を示している．単語予測のグラフは，確率の高い上位 10 位までを示しており，横軸は予測された単語の文字列を，縦軸は予測確率を示している．これらの結果より，ある程度人の感覚に即した単語が推定されていることが分かる．一方で,「女性」や「セクシー」といった単語は，上述のように学習データに多く含まれていたために推定された単語であり，必ずしも正当なプロセスで推定されているとは言えない．この問題は，シーンに含まれている物体が何であるかを考えていないことが本質的な原因であることは，前述の通りであり，これについては，物体概念を含む複数の概念を階層的に組み合わせることで解決することができると考えられる [宮澤 19].

4. 統合認知モデルと感情

上述の例は，非常に単純化したものであり情動や感情の本質を捉えているとは言い難い．しかし，例えば自然言語処理において，大量の言語資源から単語の分散表現を学習することで語彙同士の関係性から単語の意味を規定する考え方とは異なり，視覚情報が身体に及ぼした影響の知覚と語彙の結びつきに単語の意味を見出している．

この考えをさらに推し進め，情動や感情をモデル化するためにはどうすれば良いであろうか．我々は，宮澤らが提案している統合認知モデル [宮澤 19] に内受容感覚を取り入れるのが，現在実現可能な一つの方向性であると考えている．このモデルは，ロボットのマルチモーダルなセンサ情報や体勢感覚などをモダリティ毎に構造化し，さらにそれらの構造化によって得られる概念（潜在情報）同士の関係を生成モデルとして見出すものである．モデル全体は，強化学習の枠組みや時系列学習の枠組みと統合され，ボトムアップな自己組織的プロセスだけで

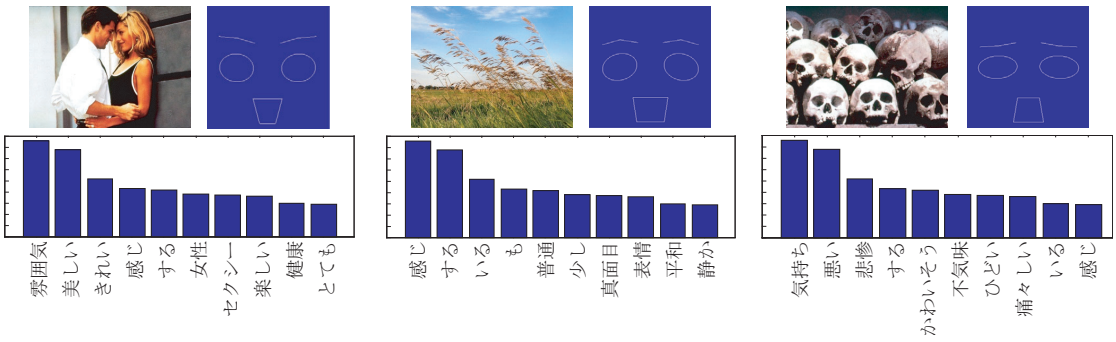

図 4: 単語予測の結果

なく，意思決定や長期の行動計画などに概念構造が影響を受ける．最も単純な考え方は，このモダリティの一つとして身体からの信号を取り入れることである．このモデルで興味深いのは，体勢感覚やモーターコマンドに基づく自他の分離（カテゴリ分類）がなされ，それが他者モデルの獲得につながる可能性があることである．そして，その他者モデルの中には，自身の内受容感覚に基づく情動状態が存在し，他者の情動状態の予測やその言語との結び付けなどが可能となる．

5. まとめ

本稿では，情動/感情を内受容感覚に基づいてモデル化することを検討した．しかしここでの議論は，内受容感覚が与えられたことを仮定してその先にどのような処理があるかを機能レベルで議論しているに過ぎない．逆に言えば，内受容感覚がある種のベクトルとして内部表現されるとすれば，その先は記号創発ロボティクスの枠組みで，知能の一部として議論できる可能性がある．したがってまず重要なことは，いかに内受容感覚を物理的なロボットで作りだすかということかもしれない．それは，バッテリー残量といったロボティクスで扱われる直感的な恒常性で成立するものなのか？尾形らの研究 [Ogata01] は，こうした点を実ロボットで検討した最初の例であり大変興味深いが，その結論はまだ十分明らかではない．また内受容感覚は，モデルからのトップダウンな信号に影響を受けるはずである．今後，内受容感覚やその予測符号化，環境や自他認知，他者モデル，本稿で述べたモデルなどを結びつけつつ，社会的な感情を含めた知能全体をモデル化することで感情のメカニズムを解明し，感情をもったロボットの実現を目指していく必要があると考えている．

謝辞

本研究の一部は，JST CREST(JPMJCR15E3) 及び科研費新学術領域「認知的インタラクションデザイン学」(26118001) の支援を受けて実施した．

参考文献

[Attamimi14] M.Attamimi, M.Fadil, K.Abe, T.Nakmura, K.Funakoshi, T.Nagai, ”Integration of Various Concepts and Grounding of Word Meanings Using Multilayerd Multimodal LDA for Sentence Generation”, IROS2014, pp.2194-2201, 2014

[Breazeal03] C.Breazeal : Emotion and sociable humanoid robots, E. Hudlika (ed), International Journal of Human Computer Interaction, 59, pp.119-155 (2003)

[Damasio03] Damasio, A.: Looking for Spinoza: Joy, Sorrow, and the Feeling Brain, Mariner Books (2003)

[Feldman17] L. Feldman-Barrett: How emotions are made: The secret life of the brain. New York, NY: Houghton Mifflin Harcourt (2017)

[JLang05] P.JLang *et al.*:International affective picture system (IAPS): Affective ratings of pictures and instruction manual, Tech. Report A-6, Univ. of Florida, Gainesville, FL (2005)

[Nakamura11] T.Nakamura, T.Araki, T.Nagai, N.Iwahashi, “Grounding of Word Meaning in Latent Dirichlet Allocation-Based Multimodal Concepts”, Adavanced Robotics 25, 2189-2206, 2011

[Ogata01] T.Ogata, S.Sugano: Emotional Communications Between Humans and the Autonomous Robot which has the Emotion Model, ICRA99, pp.462-467 (1999)

[Taniguchi16] T.Taniguchi, T.Nagai, T.Nakamura, N.Iwahashi, T.Ogata, and H.Asoh: Symbol emergence in robotics: a survey, Advanced Robotics, Vol.30, pp.706-728 (2016)

[浅田 15] 浅田: 情動発達ロボティクスによる人工共感設計に向けて, 日本ロボット学会誌, vol.32, no.89 pp.666-677 (2014)

[乾 18] 乾:感情とはそもそも何なのか:現代科学で読み解く感情のしくみと障害, ミネルヴァ書房 (2018)

[大平 17] 大平：予測的符号化・内受容感覚・感情, エモーション・スタディーズ, 第 3 巻第 1 号, pp.2 — 12 (2017)

[堀井 13] 堀井 他:乳児期の触覚優位性を利用した複数感覚情報の統合による情動文化モデル, 第 31 回ロボット学会学術講演会, RSJ2013AC1P3-04 (2013)

[宮澤 19] 宮澤, 青木, 堀井, 長井: 統合認知モデルによるロボットの概念・行動・言語の同時学習, 人工知能学会全国大会 (2019)