《第16回》内発的動機付けによるエージェントの学習と発達

浅　田　稔*

*大阪大学大学院工学研究科，大阪府吹田市山田丘 2-1
* Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan
*E-mail: asada@ams.eng.osaka-u.ac.jp

キーワード：認知発達ロボティクス (cognitive developmental robotics)，内発的動機付け (Intrinsic Motivation)，構成的手法 (synthetic approach)
JL 012/13/5201–0011 ©2013 SICE

1. はじめに

本リレー解説も第 16 回を迎え，最終回のパネルディスカッションを除けば，通常解説としては最終回である．既に，これまでの解説で強化学習の最近の発展がかなり網羅的に紹介されているので，本稿では，少し大きな視点から，強化学習の周辺の動きを探ってみようと思う．

環境に存在するエージェントがなぜ学習するかという，より根源的な課題に対して，外部からの報酬を獲得する動機付け (Extrinsic Motivation：以降，EM と略記) だけでなく，自身で内部から報酬を生成するメカニズムとして，内発的動機付け，Intrinsic Motivation (以降は IM と略記) が近年，注目を浴びている[1]．IM に関する自律エージェントの研究は，二つのコミュニティから生じており，目的も若干異なる．

(a) 能動学習のコミュニティからは，強化学習におけるサンプリングの効率を上げるための IM として提案．エージェントの世界に対する知識とそれにもとづく制御能力の最大化を支援することが目的．
(b) 発達的学習のコミュニティからは，累積的でオープンエンドの学習をするための IM として提案．ロボットが動的な環境で生涯学習し発達し続けることが目的．

前者は，通常の機械学習屋さんの立場で，学習を効率よく進めるための手段であり，すでに解説されている探索と利用のトレードオフの課題[2] も含まれる．後者は，より生物学的な意味合いでの，発達的観点から，その漸進性，拡大性，自律性，能動性に着目しており，動機付け自身は，好奇心とも強く関連し，さらには，エージェント自身の情動や認知の学習・発達にも繋がる．

本稿では，IM に関連する動向を概説するが，まず最初に，筆者らの古い仕事のロボカップを題材としたロボットの学習課題を簡単に復習し，そのなかで基本問題が扱われていることを確認する．次に，基本課題を追及する上で，より根源的な認知発達にチャレンジしている認知発達ロボティクスを紹介する．その中の一要素としての動機付けに関して，歴史的な流れを復習し，脳神経科学との関連，計算モデル及び実験等を紹介する．最後に，強化学習のみならず，それを含めた，より大きな視点での今後の課題を示し，まとめる．

2. ロボカップドメインでの強化学習

実ロボットへの応用の課題は，解説[3] でも説明されているが，ロボカップの課題に応じて，復習すると[4),5)]，

1. 学習時間：理論的には無限に学習するが，実世界ではすべてが限られている．ロボットの場合，無限の試行を繰り返すことなどできず，ロボットが摩耗し，実験の続行が困難である．人間の場合も，何度も失敗が続けば，それこそ動機を失う．そこで，易しいタスクからの学習 (LEM:Learning from Easy Missions[6]) を設定することで，理論的に，探索時間を状態行動空間のサイズの指数オーダーから線形オーダーに圧縮可能である．先験的にタスクの「易しさ」が分かれば，問題はないが，そうでない場合，その確信のなさに従い，線形オーダーから遅くなるが，元の学習の収束性が保証されていれば，同様に保証される．

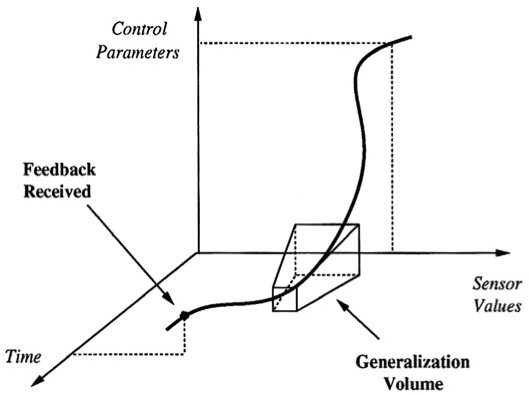


図 1　Credit Assignment Problem[7]

2. 状態行動空間：状態が格子状で行動が格子間の移動などの理想的な状態行動空間は，イベントベースの抽象的な状態行動空間を除き，実世界ではほとんどありえず，セグメンテーション課題と呼ばれる大基本問題の一つである．報酬が与えられる時間も含めて CreditAssignment Problem[7] と呼ばれている (図 1)．状態行動

空間を再帰的に定義することで，状態行動空間構成の「鶏と卵」問題を解消した手法[8] が提案されている．また．初期を一状態とし，連続の状態行動空間を線形関数近似により分割する手法[9] では，線形関数近似に加え，報酬 (ゴール到達) の成否による細分化も含まれている．最近では，ベイズ推定の枠組みで，状態・行動空間を自律的に分割する機構をもつ強化学習法が提案されており，解説[3] で紹介されている．

3. スケールアップ：より複雑なタスクへの応用として，階層構造化とマルチエージェント化の課題が挙げられる．前者では，MOSAIC[10] が有名だが，高橋ら[11] は，均一な強化学習器を多く準備し，階層のレベルを，それらの能力と環境に依存して (事前に指定しない)，自律的に構造化する手法を提案している．マルチエージェント学習では，同時学習による学習過程の不安定化(注1)が課題である．Asada et al.[12] は初期段階の交互学習をコンピュータシミュレーションで実施し，中級レベルからの同時学習で実ロボットでのスキルアップを可能にした (図 2)．さらに，内部ら[13] は，学習者の観測と行動を通して，学習者と他者の行動の関係を局所予測モデルとして推定した．局所予測モデルは線形の状態空間表現を持ち，学習者と他者の関係の複雑さは推定される状態ベクトルの次元で表現され，情報量基準をもとに決定された．

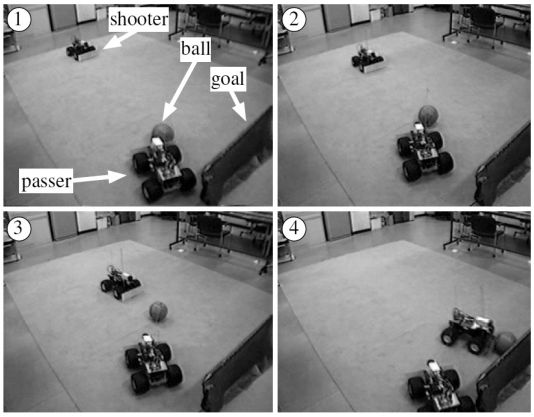


図 2　パッサーとシューターの協調行動[13]

上記の一連の強化学習関連研究を通じて，セグメンテーションの大基本問題は，状態行動空間構成課題として，行動主体に依存した形で問題が定式化された．但し，ロボットは移動台車であり，まだまだ，自由度が少数である．さらに，自己，他者などのカテゴリー化は，人間の認知課題として重要かつ未解決であり，筆者の興味は，以下で述べる認知発達ロボティクスに移って行った．

(注1) 初心者二人のテニスを連想するとよい．どちらも下手なので，練習もできない．相手がコーチだと定まったボールを初心者に呈示するので，初心者は安心して練習できる．

3.　認知発達ロボティクスのアプローチ

認知発達ロボティクス[14),15)] とは，理解の対象となる人間の発達モデルを人工物の中に埋め込み，環境の中で作動させ，その挙動から，発達モデルの新たな理解を目指すものである．その核となるアイディアは，物理的身体が他者を含む環境との相互作用から情報を構造化するための「身体性」と「社会性」であり，それをシームレスに繋ぐのが「認知発達」である．その設計論は，身体を通じて世界に働きかけるロボット内部の知覚行動設計，およびロボットが上手に学習や発達できるような環境，特に教示者をはじめとする他者の行動を含む環境設計である．そこには，「適応性」と「自律性」も含まれる．

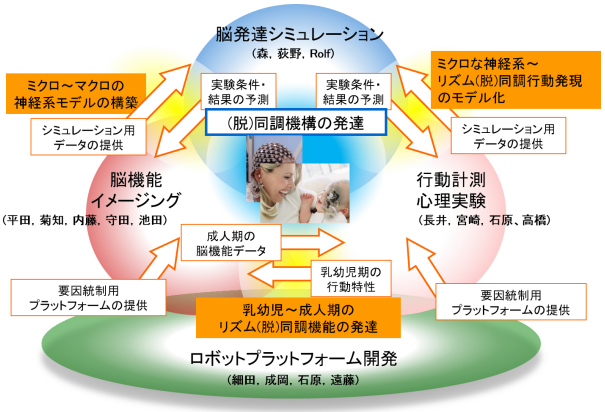


図 3　神経ダイナミクスから社会的相互作用へ至る過程の理解と構築による構成的発達科学プロジェクトの概要

具体例は，筆者らが手がけた JST ERATO 浅田共創知能システムプロジェクト(注2) である．共創の意味は二重で，一つは，他者を含む環境との相互作用を通じた知能の創発，二つ目は，単一の学問分野ではなく，学際的な協働の意味を込めている．胎児シミュレーションから始まり，人工筋による動的運動学習，自他認知，さらに社会的相互作用として音声模倣，共感発達，コミュニケーション発達に及び，広範な課題を扱ってきた．現在では，「神経ダイナミクスから社会的相互作用へ至る過程の理解と構築による構成的発達科学 (科研特推 2012-2016)」に引き継がれ，赤ちゃんが外界との相互作用を通じて，自己と非自己から，自己に似た他者 (養育者)，自己と異なる存在の認知の発達過程を説明可能な計算モデルの提唱とそのイメージングや心理・行動実験による検証を通じたモデル精緻化を目指している．この過程で，自他認知のキーとなるミラーニューロンシステム[16] が構築・発達していくと想定される[17]．詳細はウェブ頁(注3) に譲るとして，全体の簡単な概要を図 3 に示す．計

(注2) 2005-2011: www.jeap.jp 参照
(注3) http://www.er.ams.eng.osaka-u.ac.jp/の「プロジェクト」参照

算モデル，イメージング，心理・行動実験，ロボットプラットフォームの各グループが密に結合して，自他認知の発達原理の解明を目指している．

これらの一連の研究のなかで，エージェントの行動の駆動原理としての情動(注4)があると想定される．それは好奇心や動機などのエージェントの内的な属性と結びつき，様々な行動を生成する．以下では，この動機付けについて，概説する．

4. 心理学的視点からのIM

「はじめに」で述べたように，内発的動機付けのIMについては，強化学習と発達ロボティクスの両方のコミュニティでホットな話題ではあるが，その定義や計算モデルについて，明確なコンセンサスが得られている訳ではない．ここでは，文献[18),19)]などに従って，定式化を試みてみよう．

4.1 それ自身のために探究する活動

Ryan and Deci[20)]に従えば，

- IM：行為それ自身が本質的にもつ楽しみや満足のための動機．興味，挑戦など
- EM：行為自身とは別の結果を得ることが目的の行為をとり続ける動機．操作的価値 (instrumental value)

IM行動は，乳幼児の時期から観られ，新しく出くわした物を握ったり，投げたり，噛んだり，物に対して叫んだりなどである．余談だが，米国の統計で2001年生まれの子どもの死亡要因のトップは246項目中で転落事故であり，これは交通事故のリスクと等価と言われている[21)]．つまり，危険を介せず，歩行などにチャレンジしている様で，まさにIMの極致であろう．大人になっても，クロスワードパズルを解いたり，絵を描いたり，庭いじりなどに興ずる．英語のintrinsic, internal, extrinsic, externalの違いは混同の要因であるが，Oudeyer and Kaplan[18)]は，以下のように説明する．宿題をする子どもを想定すると，

a 宿題をしないと親から制裁を受けるので，その制裁を回避する場合．行動要因は外的 (external) で，目的が親からの制裁回避であり，内発的ではなく，外発的 (extrinsic) である．

b 大人になったときに自身の夢の職業につきたいと願っての場合，行動要因は内的 (internal) であるが，目的は良い就職に繋がることなので，外発的 (extrinsic) である．

c 宿題が面白くて，それ自身のためにやる場合．新しい知識獲得への好奇心で，ビデオゲームを楽しむようにするので，行動要因は内的 (internal) であり，目的がそれ自身なので，内発的 (intrinsic) である．

これらは，個別ではなく，同時に存在する場合もある．

4.2 能動的に内発的動機付けするものは何か？

1. 動因理論 (Theory of drives, 1950年前後)：飢えや痛みなどの不利な状況を軽減する操り動因，探索動因などでIMや探索活動の説明を試みたが，不調．
2. 認知的不協和軽減 (reduction of cognitive dissonance) 理論 (1950年代後半)：内的認知構造と現在知覚される状況との差異を軽減することがIMの駆動原理と主張．しかし，人間の行動は不確実性を増大する意図の行動もあるので，これも不調．
3. 最適不一致 (optimal incongruity) 理論 (1960年代半ば)：知覚と刺激の差異が興味ある対象．最も報酬があるのは，新規性が半ば，すなわち既知と完全な新規の間．
4. イフェクタンス，個人的原因帰属 (personal causation(注5)) 有能さと自己決定 (competence and self-determination) のための動機付け：他者や物体，そして自身の行動の制御自由度に依存．換言すれば，効果的相互作用の量が動機付けを形作る．

4.3 Collative variables

前節の3,4あたりをまとめたのが，Berlyne(注6)で，簡単に言えば，新規性，変化，驚き，不一致，複雑，曖昧などで参照される刺激によって起動されるもので，これらをまとめて，Collative variablesと呼んだ(注7)．

上記を計算論の立場から明らかにするために，情報理論の測度を使う．典型的なIMは，驚きを探し出す動機である．非IMの典型は，自身の身体の恒常性を維持するために食料や水を探そうとする動機である．

5. IMの神経科学的基盤

強化学習に関連する神経基盤，とくに神経修飾物質と意志決定については，解説[22)]を参照して頂くとして，ここでは，以降の計算モデルと関連する三種類のIM関連の神経基盤について簡単に触れる[23)]．

5.1 上丘・ドーパミン・大脳基底核系

一過性の(注8)ドーパミン作働性信号の機能に関する理論[24)]が提案されており，この理論によれば，期待していないイベントが上丘を活性化し，それがドーパミン領域を活性化し，一過性のドーパミンのバーストを引き起こし，大脳基底核に達する．ドーパミンは，大脳基底核に表れたイベントを引き起こした行動とそのコンテキストの全情報の連関を引き起こすスタンプの役目を担っている．この現象は，スキル学習に繋がるが，繰り返しが多くなると上丘

(注4) 認知科学 (cognitive science) に対して，感情科学 (affective science) が提唱されているが，認知発達ロボティクスでは，広義の認知発達過程を扱い，その中には，情動や感情なども必然的に含む．

(注5) 社会心理学では自己原因性とも訳されている．

(注6) 執筆時，原著を入手できなかったので，引用文献[18)]の中の抜粋のみからの摘要

(注7) 訳語がみあたらないので，原文そのままとする

(注8) phasicの訳

は学習信号を抑制し，スキル遂行が止む．また，必要に応じて呼び出される．これらの機構は IM の典型的特徴を持つ．すなわち，スキル及びその帰結に関する知識獲得，学習信号の脳内の上丘で生成，そして，学習後の学習信号の停止である．

5.2 海馬・ドーパミン系

二番目は，海馬・ドーパミンと記憶生成に関するもの[25]である．海馬は，未知の物体，既知物体の未知配置，既知物体の未知系列などの知覚により，活性化する．そして，海馬は腹側被蓋領域 (VTA) のドーパミン作動性ニューロンを活性化し，これが海馬と VTA のターゲット領域の前頭皮質の間での新しい記憶の生成に繋がる[26]．このことは，海馬と皮質が新規の刺激-刺激の連想学習を促す．これらの機構は以下の特徴を持つ．すなわち，新規の物体と連想学習による知識獲得，学習信号の脳内の海馬で生起，知覚された物体が記憶されたことによる学習信号停止などである．

5.3 神経修飾物質：ノルアドレナリン，アセチルコリン

TD 誤差とドーパミンに関する知見に関しては，解説[22]に詳しいが，ここでは，ノルアドレナリンとアセチルコリンに関連する提案を示す[27]．脳は，世界の機能について，トップダウンの期待を持っており，これらの期待が，より正確な行動を生成するために，ノイズにまみれた入力と統合される．期待は，実際の経験とのミスマッチに基づいて獲得される．最初のキーアイデアは，アセチルコリンのレベルは高い不確実性，すなわち，期待がそんなに信頼できないと想定されたときに高いことで，これは．アセチルコリンが，どの程度期待もしくは実際の知覚に頼るべきか，どれくらい集中して期待を更新するかの両方を調整している事を意味する．二つ目は，ノルアドレナリンが，信頼できると想定された期待と直接知覚のミスマッチがあったときに期待されなかった不確実性信号を出すことである．これらの機構は以下の特徴を持つ．すなわち，ノルアドレナリンとアセチルコリンの信号は，世界に対する知識獲得を支援する，学習信号は脳内の皮質で生起する，そして，世界が期待通りに動けば，学習信号は徐々に消失することである．

6. IM の計算モデル

IM の計算モデルは，当然のことながら，強化学習モデルが基本であるが，それを明確にするために，従来の強化学習のアーキテクチュアを，IM を想定した場合に拡張したアーキテクチュアの概念図が示されている (図 4)．これは，Andrew Barto が，ICDL-EpiRob2011 で行った招待講演のスライドから再現したもので，エージェント自身の中に Critic が内包され，内発的動機に基づく報酬を生成している．

具体的な例として，前節の神経モデルを参考しながら，三つの計算モデルを簡単に紹介する．

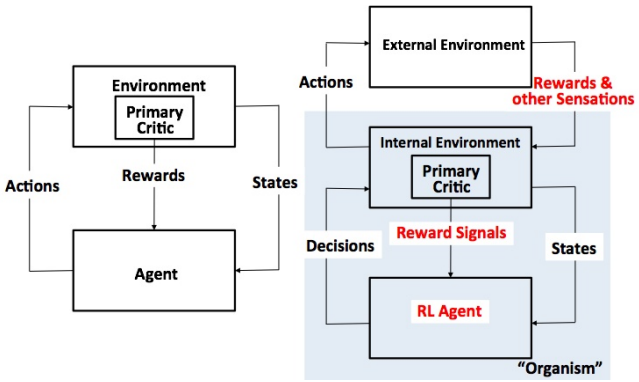


図 4 強化学習のアーキテクチュアの見直し (Andrew Barto の招待講演 (ICDL-EpiRob2011) スライドから再現)

6.1 知識ベースの IM モデル

予測誤差 (神経修飾物質) や新規性 (海馬・ドーパミン系) が関連し，それらにより知識獲得を基本とするモデルである．情報の理論と分散モデル，及び予測モデルに分かれる．前者では，状態，任意の状態遷移，条件付き状態遷移の確率を総じて，$P(e)$ として表し，分割空間のための分布関数の形状特徴の測度を以下のエントロピーで表す．

$$H(E) = -\sum_{e \in E} P(e) ln P(e) \qquad (1)$$

新規性への引き込みは IM の典型であり，直接的な表現は，実際に観測された e に対して，その観測確率が小さいとき大きな報酬 $r(e)$ となる以下が定義されている．

$$r(e) = C(1 - P(e,t)) \qquad (2)$$

ここで，C は定数．これは不確実性動機 (UM: Uncertainty motivation) と呼ばれ，ロボットの視点制御の設計などに利用されている[28]．

エンパワーメント (Empowerment) と呼ばれる報酬測度が提案されている[29]．これは，センサー情報を最大化する行動系列を生成するようにエージェントを奨励する．任意のステップ数の間隔で，行動系列 $A_t, A_{t+1}, ..., A_{t+n-1}$ から，知覚 S_{t+n} へのチャンネル容量として，以下のように定義される．

$$r(A_t, A_{t+1}, ..., A_{t+n-1} \rightarrow S_{t+n}) = max_{p(\vec{a})} I(A_t, A_{t+1}, ..., A_{t+n-1}, S_{t+n})$$

ここで，$p(\vec{a})$ は，行動系列 $\vec{a} = (a_t, a_{t+1}, .., a_{t+n-1})$ の確率分布関数，I は相互情報量を表している．Capdepuy et al.[29] は，マルチエージェント環境での探索問題に適用し複雑な行動系を引き出している．

予測モデルも IM の典型であろう．ロボットの知識や期待は完全な確率分布で表現されるとは限らず，ニューラルネットワークや SVM などの予測器で表される．これらの

予測器を Π とし，状態やその属性の一般的な標記を先にならい e とし，現在の感覚運動コンテキスト，そして可能ならば過去のコンテキストも符号化する構造を $SM(\rightarrow t)$ と表すと，

$$\Pi(SM(\rightarrow t)) = \check{e}(t+1) \quad (3)$$

ここで，$\check{e}(t+1)$ は，予測されたイベントで，実際のイベント $e(t+1)$ との差を誤差 $E_r(t)$ として定義する．

$$E_r(t) = ||\check{e}(t+1) - e(t+1)|| \quad (4)$$

予測新規動機 (Prediction novelty motivation: NM) は，予測誤差が最大の時，最大報酬になる．

$$r(SM(\rightarrow t)) = CE_r(t) \quad (5)$$

ここで，C は定数．Barto et al.[30] は，再利用可能なスキルの階層構造化とその拡張に，この誤差規範を用いた．

学習進度動機 (Learning progress motivation: LPM) の最初の提案は，Schmidhuber[31] が行った．学習の進み具合の測度として，予測器 Π の予測誤差の差を用いた．すなわち，同じ感覚運動コンテキスト $SM(\rightarrow t)$ の下で，最初の予測と学習則によって更新された予測器 Π' による予測との差で表される．つまり，

$$r(SM(\rightarrow t)) = E_r(t) - E'_r(t) \quad (6)$$

ここで，

$$E'_r(t) = ||\Pi'(SM(\rightarrow t)) - e(t+1)|| \quad (7)$$

である．これにより，ゴール指向のオンライン学習向けの適応的世界モデルを用いた「好奇心神経制御器 (curiosity neural controller)」を定式化した．

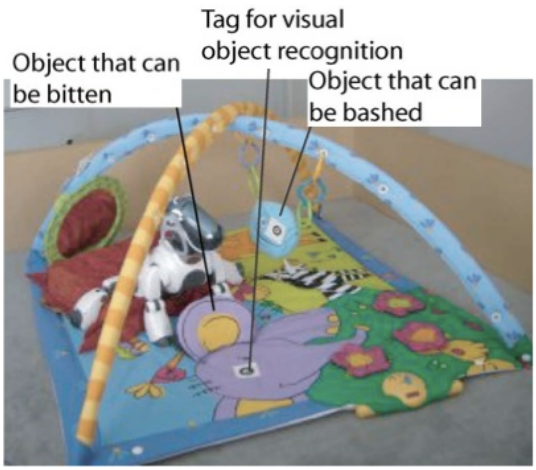


図 5　感覚運動アフォーダンスの実験環境[19]

学習の進み具合を測る他の手法として，Oudeyer et al.[19] は，類似状況を領域に統合し，その中で比較する手法を提案している．領域数やその境界は，適応的に更新される．彼らは，これを AIBO を用いた感覚運動アフォーダンス実験にもちいた (図 5)．

予測新規動機の変形として，予測既知動機 (Predictive familiarity motivation: FM) がある．予測可能かつ既知の場合を好む場合で，

$$r(SM(\rightarrow t)) = \frac{C}{E_r(t)} \quad (8)$$

と表され，Andry et al.[32] が，アーム付きの移動台車が視覚を通じて，感覚運動の不変項を学習するケースに応用した．

明示的な IM モデルとは称していないが，顕著性に基づく好奇心により，語彙爆発のモデルを菊池ら[33] が提案し，シミュレーションと実ロボットの実験で，その有効性を示している．

6.2　能力ベースの IM モデル

自身の決断能力に依存した動機による行動学習で，ゴール達成と期待される時刻 t_g に期待されるゴール g を自分で設定したとしよう．この能力の測度として，

$$l_a(g, t_g) = ||\overline{g(t_g)} - g(t_g)|| \quad (9)$$

が定義される．ここで，$\overline{g(t_g)}$ は，予測されたゴールであり，$g(t_g)$ は，実際の状態もしくはイベントを指す．次の時刻 $t_g + 1$ のゴールは，将来のこの測度の累積を最大化する方向に行動選択する，すなわち，常々，果敢に未知領域に挑戦する戦略に対応する．これは，無能力最大化動機 (Maximizing incompetence motivation: IM) と呼ばれ，以下で定義される．

$$r(SM(\rightarrow t), g, t_g) = l_a(g, t_g) \quad (10)$$

このモデルが，怪我して痛い目に，さらには死にたる危険性も顧みずチャレンジする乳幼児のケース[21] に対応しそうである．もちろん，逆の最大能力進度 (Maximizing competence progress) もありうる．

知識ベースと能力ベースの両方を含みうるケースとして，乳幼児の発声発達の自己組織化にならった音韻学習のアプローチがある[34]．好奇心駆動型の学習 (知識ベースの IM モデル) に基づき，発声域の行動計画の自己組織化がなされ，能力ベースの IM モデルとも見なせる．

6.3　形態学的 IM モデル

上記二つの IM モデルは，現在の状況と過去や記憶の状況との比較の情報表現に依存するが，形態学的 IM モデルは，同時刻に知覚される刺激間の比較に依存する．代表的なものは，同期性動機 (Synchronicity motivation: SyncM) で，同期性による報酬最大化が図られる．これは，因果関係・随伴関係を学習する際の同期性動機にもとづく同期現象発見が代表例であろう．同期すること自体に興味あるので，一旦発見された同期性は，それを繰り返し知覚するための予測可能な行動の帰結として，同期性が観測される．

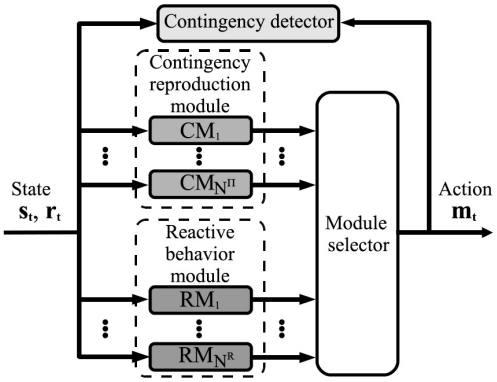


図 6　随伴性の発見と再生アーキテクチュア[35)]

Sumioka et al.[35)] は，これにより，一旦発見された同期による随伴関係の再帰的構造化を共同注意関連の行動学習で示している．図 6 に，そのアーキテクチュアを示す．

7．IM の今後の課題

心理学の知見で示されているように，IM と EM は独立ではなく，相互に関連し，同時に存在する可能性や，時期を経て，発達的変化をする場合もある．ここでは，情報量基準を用いて，IM の計算モデルを紹介してきたが，個々のモデル間の遷移や，社会性を陽に取り込んだモデルにはなっていない．社会性による動機付けは EM と取られがちだが，必ずしもそうではない．例えば，発声における相互模倣の場合[36)]，模倣されることの喜びが，発声の模倣行為の動機付けになっているので，互いの IM が環境を介して相互作用していると見なせる．人間の死亡要因のメタ解析[37)]では，喫煙，アルコール，大気汚染などの直接的要因を抑えて，社会的関係性が上位を占めた．それほどに人間の場合，社会的関係性が重要である．Ogino et al.[38)] は，社会的関係性を要求する乳幼児の様子を表す Still Face パラダイム[39)] (注9) において，その計算モデルを構築し，社会的関係性に対する要求のメカニズムを学習を通じて表している．図 7 に学習結果を示す．左が Still Face 期間を含む相互作用時間帯における養育者の推定された情動状態で，右が学習者 (9ヶ月時想定) の社会的関係性尺度で，Still Face 期間に落ちている様子が窺える．これらも含めて，今後，社会性の明示的な IM モデルへの導入，ならびに発達的変化も考慮したモデルへの拡張や，それらのロボットへの実装による検証などが望まれる．

(注9) 乳幼児と親との相互作用中に突然，親が乳幼児からの応答に何も反応しない静止顔になると，乳幼児の笑顔が減少し，ぐずり，親の注意を引こうと発声することから，他者との関係性を維持したいという欲求が親子間相互作用を動機づけていると考えられているパラダイム．ただし，メカニズムの詳細は不明とされている．

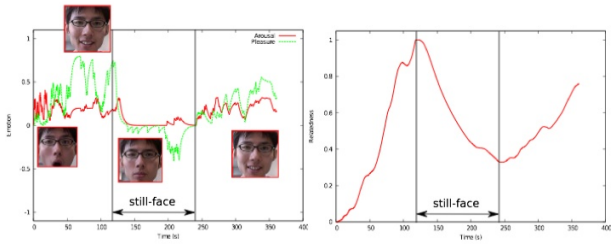


図 7　社会的関係性を動機付けとする学習結果[38)]

8．おわりに

IM に関連する心理学における歴史的考察，脳神経科学的知見，計算モデルを紹介した．IM 自身は，認知発達ロボティクスの課題でもあり，社会性を考慮した IM は，人間の行動の駆動源であり，認知発達過程において重要な位置を占める．ロボットの行動原理設計の観点からも重要な研究項目である．今後はより広範な視点から，より精緻な計算モデルの進展が望まれ，それが，自己や他者の認知を含めた，人間の心的機能の構成的理解に繋がると期待される．

(2013 年 9 月 10 日受付)

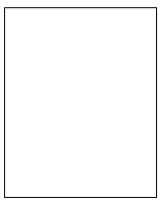
参　考　文　献

1) Manuel Lopes and Pierre-Yves Oudeyer. Guest editorial active learning and intrinsically motivated exploration in robots: Advances and challenges. *IEEE Transactions on Autonomous Mental Development*, Vol. 2, No. 2, pp. 65–69, 2010.
2) 牧野貴樹. 強化学習の最近の発展《第 2 回》探索と利用のトレードオフとベイズ環境モデル. 計測と制御, Vol. 52, No. 2, pp. 154–161, February 2013.
3) 保田俊行, 大倉和博. 強化学習の最近の発展《第 8 回》連続空間における強化学習によるマルチロボットシステムの協調行動獲得. 計測と制御, Vol. 52, No. 7, pp. 648–655, July 2013.
4) 浅田稔. 実環境におけるロボットの学習・進化的手法の適用と課題. 計測と制御, Vol. 38, No. 10, pp. 650–653, 1999.
5) 浅田稔編著. RoboCupSoccoer ロボットの行動学習・発達・進化. 共立出版, 2002.
6) Minoru Asada, Shoichi Noda, Sukoya Tawaratumida, and Koh Hosoda. Purposive behavior acquisition for a real robot by vision-based reinforcement learning. *Machine Learning*, Vol. 23, pp. 279–303, 1996.
7) J. H. Connel and S. Mahadevan. “Rapid task learning for real robot”. In J. H. Connel and S. Mahadevan, editors, *Robot Learning*, chapter 5. Kluwer Academic Publishers, 1993.
8) Minoru Asada, Shoichi Noda, and Koh Hosoda. Action based sensor space segmentation for soccer robot learning. *Applied Artificial Intelligence*, Vol. 12, No. 2-3, pp. 149–164, 1998.
9) 高橋泰岳, 浅田稔. 実ロボットによる行動学習のための状態空間の漸次的構成. 日本ロボット学会誌, Vol. 17, No. 1, pp. 118–124, 1999.
10) 川人光男, 銅谷賢治, 春野雅彦. 多重順逆対モデル（モザイク）その情報処理と可能性. 科学, Vol. 70, pp. 1009–1018, 2000.
11) 高橋泰岳, 浅田稔. 複数の学習器の階層的構築による行動獲得. 日本ロボット学会誌, Vol. 18, No. 7, pp. 1040–1046, 2000.

12) Minoru Asada, Eiji Uchibe, and Koh Hosoda. Cooperative behavior acquisition for mobile robots in dynamically changing real worlds via vision-based reinforcement learning and development. *Artificial Intelligence*, Vol. 110, pp. 275–292, 1999.
13) 内部英治, 浅田稔, 細田耕. 複数の学習するロボットの存在する環境における協調行動獲得のための状態空間の構成. 日本ロボット学会誌, Vol. 20, No. 3, pp. 281–289, 2002.
14) 浅田稔. 身体・脳・心の理解と設計を目指す認知発達ロボティクス. 計測と制御, Vol. 48, No. 1, pp. 11–20, Jan 2009.
15) Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 1, pp. 12–34, 2009.
16) ジャコモ・リゾラッティ(著), コラド・シニガリア (著), 茂木健一郎 (監修), 柴田裕之 (翻訳). ミラーニューロン. 紀伊国屋書店, 2009.
17) 浅田稔. 共創知能を超えて-認知発達ロボティクスよる構成的発達科学の提唱-. 人工知能学会誌, Vol. 27, No. 1, pp. 2–9, January 2012.
18) P-Y. Oudeyer and F. Kaplan. How can we define intrinsic motivation? In *Proceeding of the 8th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems (Epirob 2008)*, pp. 93–101, 2008.
19) P-Y Oudeyer, F. Kaplan, and V.V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 2, pp. 265–286, 2007.
20) Richard M. Ryan and Edward L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, Vol. 25, No. 1, pp. 54–67, 2000.
21) Amy S. Joh and Karen E. Adolph. Learning from falling. *Child Development*, Vol. 77, No. 1, pp. 89–102, 2007.
22) 吉本潤一郎, 伊藤真, 銅谷賢治. 強化学習の最近の発展《第 10 回》脳の意思決定機構と強化学習. 計測と制御, Vol. 52, No. 8, pp. 749–754, August 2013.
23) Gianluca Baldassarre. What are intrinsic motivations? a biological perspective. In *IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2011)*, pp. CD–ROM, 2011.
24) P. Redgrave and K. Gurney. The short-latency dopamine signal: a role in discovering novel actions? *Nature reviews Neuroscience*, Vol. 7, No. 12, pp. 967–975, 2006.
25) D. Kumaran and E. A. Maguire. Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, Vol. 17, No. 9, pp. 735–748, 2007.
26) J. E. Lisman and A. A. Grace. The hippocampal-vta loop: controlling the entry of information into long-term memory. *Neuron*, Vol. 46, No. 5, pp. 703–713, 2005.
27) A. Yu and P. Dayan. Expected and unexpected uncertainty: Ach and ne in the neocortex. In *in Advances in Neural Information Processing Systems 15 (NIPS)*, pp. 157–164M, 2002.
28) Xiao Huang and John Weng. Novelty and reinforcement learning in the value system of developmental robots. In *In Proceedings of the 2nd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems*, pp. 74–55, 2002.
29) Philippe Capdepuy, Daniel Polani, and Chrystopher L. Nehaniv. Maximization of potential information flow as a universal utility for collective behaviour. In *In Proceedings of the 2007 IEEE Symposium on Artificial Life*, pp. 207–213, 2007.
30) Andrew G. Barto, Satinder Singh, and Nuttapong Chentanez. Intrinsically motivated learning of hierarchical collections of skills. In *In Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, 2004.
31) J. Schmidhuber. Curious model-building control systems. In *In Proceeding International Joint Conference on Neural Networks, volume 2*, pp. 1458–1463, 1991.
32) Pierre Andry, Philippe Gaussier, Jacqueline Nadel, and Beat Hirsbrunner. Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms. *Adaptive behavior*, Vol. 12, No. 2, pp. 117–138, 2004.
33) 菊地匡晃, 荻野正樹, 浅田稔. 顕著性に基づくロボットの能動的語彙獲得. 日本ロボット学会誌, Vol. 26, No. 3, pp. 45–54, 2008.
34) Clement Moulin-Frier and Pierre-Yves Oudeyer. Curiosity-driven phonetic learning. In *IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2012)*, pp. CD–ROM, 2012.
35) Hidenobu Sumioka, Yuichiro Yoshikawa, and Minoru Asada. Reproducing interaction contingency toward open-ended development of social actions: Case study on joint attention. *IEEE Transactions on Autonomous Mental Development*, Vol. 2, No. 1, pp. 40–50, 2010.
36) H. Ishihara, Y. Yoshikawa, K. Miura, and M. Asada. How caregiver's anticipation shapes infant's vowel through mutual imitation. *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 4, pp. 217–225, 2009.
37) Julianne Holt-Lunstad, Timothy B. Smith, and J. Bradley Layton. Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine*, Vol. 7, No. 7, p. e1000316, 2010.
38) Masaki Ogino, Akihiko Nishikawa, and Minoru Asada. A motivation model for interaction between parent and child based on the need for relatedness. *Frontiers in Cognitive Science (to appear)*.
39) Edward Tronick, Heidelise Als, Lauren Adamson, Susan Wise, and T. Berry Brazelton. The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child & Adolescent Psychiatry*, Vol. 17, No. 1, pp. 1–13, 1978.

［著 者 紹 介］

浅 田 稔(あさだ みのる) 君(正会員)



1953 年 10 月 1 日生. 82 年大阪大学大学院基礎工学研究科後期課程修了. 1995 年大阪大学教授. 1997 年大阪大学大学院工学研究科知能・機能創成工学専攻教授. 工学博士 (大阪大学) . 05 年から JST ERATO 浅田共創知能プロジェクト総括. 知能ロボットの研究に従事. ロボカップ国際委員会元プレジデント (2002 − 2008), 現理事. 2005 年から IEEE Fellow.