2A1-05

強化学習とロボットの知能

- あめとむちで知能は作れるか？ -

Reinforcement Learning and Robot Intelligence

- Can Intelligence be Realized by Carrot-and-Stick ? -

柴田 克成
Katsunari Shibata

大分大学 工学部 電気電子工学科
Dept. of Electrical & Electronic Engineering, Faculty of Engineering, Oita Univ.

1. 評価と行動の自律学習と知能

ロボットを人間並みに知能化することは、ロボット研究者の究極的な目標の一つであろう。さまざまなロボットが開発されている現在でも、人間とロボットの知能には、依然として大きなギャップがあると言わざるを得ない。

少し、具体的に考えてみよう。たとえば、動いている物体を捕獲する問題を考えた場合、捕獲する物体やその動きの認識方法、認識結果から物体の動きを予測する方法、予測された位置へ行くための制御方法などのプログラムが作成される。このとき、なぜこの物体を捕獲すべきか、どうして物体の動きを予測する必要があるのか、どうしてそういう制御法で良いのかといった問題は、人間にとっては当たり前であったり、与えなくてもロボットは動作するため、通常、ロボットに知識として与えることはない。しかし、これらのことまで考えて行動する方が、より賢いと考えることは自然である。また、ボールを見ると生物ではなく、ハエや蚊を見ると生物だと答えるプログラムを作成することはできても、精巧なハエの模型を初めて見て生物ではないと答えることは困難だろう。われわれ人間は、最初はハエや蚊が生物であるかどうかすらわからないが、動き方、近づいたときの反応を見て、時には確かめるための行動を起こすなどして、最終的にハエの模型も生物でないことを学ぶのであろう。そして、さらに、そういう経験を通して生物とは何かを積み上げていくと考えられる。

従来は、いかに人間がロボットに有効な行動をするプログラムを与えるかが問われてきた。ここで本当に賢いのは、忠実に作業をこなすロボットではなく、設計者としての人間である。しかし、設計者があらゆることを考慮して予めプログラミングしようとしても、結局フレーム問題にぶち当たってしまう。この限界を打破するためには、ロボット自身を賢くするしかない。そして、そのためには、ロボットに知識を与えるのではなく、ロボットが自ら学んで、考えて、行動する**「自律学習型」知能ロボットに大きく方向を転換するしかない**と考える。特に、行動を学習するだけでなく、そのために必要な、行動をいかに評価するか自体を自分で学習することが重要であろう。これが、前述の「なぜ」「どうして」ということを考えることにつながると期待できる。また、自律学習では、予め与えられる知識は拘束となりうるので、獲得される機能の柔軟性から、むしろ与えられる知識は少ない方が良い。以上より、**強化学習は**、外界との行動－知覚－行動のフィードバックループを有し、試行錯誤をしながら自ら評価と行動を学習することができ、さらに、報酬や罰という少ない情報から学習するため、**自律学習を実現する理想的な手段である**と言える。

連絡先：〒870-1192 大分市大字旦野原 700 番地
大分大学工学部電気電子工学科, shibata@cc.oita-u.ac.jp
http://shws..cc.oita-u.ac.jp/~shibata/home-j.html

2. 統合的学習と知能

前述の移動物体の捕獲問題の例からもわかるように、従来の知能ロボットでは、まず、ロボット全体の機能を、画像処理、画像認識、音声認識、注意、記憶、プランニング、軌道生成、制御、音声合成といった機能モジュールに分割する。そして、それぞれのモジュールの高機能化を図り、最終的に、それらをくっつけることによって全体としての高機能化を行ってきた。また、さらなる知能化のために、高次機能用のモジュールを設けるといったアプローチもとられる。しかし、これらをばらばらに開発することで本当に知能が生まれるのであろうか。

自律学習では、前述のハエの模型の例のように、予め想定していないようなことにも対応できるための非常に柔軟な学習が求められる。したがって、そこで何を入力として何をするのかは状況に応じて大きく変化させるべきであり、予め決めるべきではない。モジュール化しようとすると、結局入出力を定めたりしなければならず、柔軟性を大きく阻害することになってしまうし、モジュール間の調和的な学習も困難になる。これらのことから、**知能の実現のためには、「自律学習」とともに、センサからモータまでの「統合的な学習」が必要条件**となると筆者は考える。

3. 「強化学習は行動の学習」ではもったいない

強化学習は、行動のプランニングのための学習であると一般的に捉えられている。行動のプランニングは、分割された機能モジュールの一つであり、これを頑張って学習させて、状態空間と行動空間の間の適切なマッピングを作り上げたところで、知能と呼べるような代物にはならないのではないだろうか。これでは非常にもったいないというのが筆者の考えである。前述のように、知能の実現には、センサからモータまでの機能の統合的学習が必要であり、そのためにも強化学習が有効であると考えるのである。

センサからモータまでの間には、前述のように**さまざまな機能**が求められる。これらは、最終的に、より良い行動を行って報酬を得るために行われるものであり、それ自身も**行動の一種**であると極論することができる。つまり、より良い認識をすることでより報酬にありつけるようになり、必要な記憶をしておけば、やはり報酬にありつけるようになるはずである。逆に言えば、**与えられた目的を達成する**

ための認識、記憶などが必要なのであれば、強化学習で学習することが望ましい。前述の生物かどうかの区別に関しても、その区別をすることが、報酬を獲得し、害を逃れるために有用であるために発現してくると考えるのである。

筆者らは、現在まで、このようなさまざまな機能が本当に強化学習だけで学習できるのかということを追求してきた。そして、認識とセンサ動作[柴田 01-1]、注意と記憶と連想[Shibata 02]、異種センサ信号の統合(手先のリーチング運動、hand-eye coordination)[柴田 01-2]、予測と先回り行動[前原 02]、コミュニケーションと交渉[柴田 99]、社会行動における個性と社会性[Shibata 00-1]、報酬の分配[Shibata 00-2]などの機能について、簡単な例題を与え、強化学習(報酬や罰)のみによって学習できることを確認してきた。

また、最終的には、前述のように、これらの機能を統合的に学習することが知能に結びつくと考えている。そのためには、**ニューラルネットの使用**が非常に有効であると考える。バックプロパゲーション(BP)法を用いれば、サンプルの入出力関係を実現するために、ニューラルネット内で自律的な役割分担ができる。したがって、ロボットのセンサからモータまでをニューラルネットで構成し、強化学習に基づいて生成された教師信号で学習させれば、ニューラルネット内で**さまざまな機能が必要に応じて自律的に発現**するようになると期待される。そのような流れで Direct-Vision-Based 強化学習を提案し、物体の捕獲タスクを強化学習で学習するだけで、視覚センサ信号を入力とするニューラルネットの中間層に、ロボットが存在する空間が表現されるようになることを示した。さらに、個々のセンサセルが一般的に局所的な受容野を持つことが、学習の高速化・安定化につながることも示した[柴田 01-3]。また、最近、実機(Khepera)でも視覚センサ上の黒い物体に近づく動作を、全くの０から学習できることを確認した[飯田 02]。

4. 課題と今後の方向性

前述のように、これまでさまざまな機能が強化学習によって獲得可能であることを示してきた。しかし、**論理的思考**、**シンボルの発現や処理**に関する機能の獲得はまだ確認できていない。特に、ニューラルネットはこれらの扱いを苦手としている。したがって、シンボルの発現や処理ができることを示さない限り、ニューラルネットによってさまざまな機能の統合的学習ができ、知能につながると結論づけることはできない。連想記憶における固定点収束のダイナミクス、ノイズの影響などがシンボルの発現につながるのではと期待しているが、今しばらくの検討が必要である。

試行錯誤に基づく強化学習では、学習時間の遅さが重大な問題点となる。前述の論理でいくと、０から学習させることが重要であり、どうしても学習時間の問題にぶつかってしまう。しかし、知識を与える場合でも、たとえば ニューラルネットの**初期値という形で知識を与える**のであれば、その後の学習によって更新することができ、柔軟性の阻害にはつながらないと考える。ただし、どのような初期値からでもある程度学習できるような能力は必要であろう。

さらに、学習を加速させるためには、**身体の成長**という考え方が重要であると考えられる。われわれ人間や生物の身体は、生まれたばかりは無力な状態でいろいろな試行錯誤をし、次第に大きく成長していく。これは、行動学習時にまわりに害を与えないという働きと同時に、筋の同時活性などと同様、学習初期の探索範囲の減少などの効果から、学習の加速につながるのではないかと考えている。

また、さまざまな機能の統合的学習のためにニューラルネットが有効であることを前に述べた。しかし、ニューラルネットの構造は、通常３層か４層程度であり、人間の脳における複雑な構造とは大きく隔たりがある。これで、人間のような知能を実現させるというのはおこがましい。われわれ生物は、生後すぐにニューロン間のシナプス数が大幅に増大し[津本 86]、また、化学物質によってニューロンの軸索が伸長していると言われている[畠中 92]。このようなことから、**ニューロンが成長**することで、ニューラルネットが**必要に応じて自律的かつ柔軟に構造を獲得する**ことが必要ではないかと考えている。そして、反射のような低次の機能から、学習に伴うニューラルネットの成長とともに、高次の機能が獲得できるようになればと期待している。

最後に、本パネルのテーマの一つである「強化学習(知能)」研究における異分野の交流に関連して考えることを述べる。筆者は、本研究の目的が「**脳のモデル**」なのかそれとも「**工学的な応用**」なのかということを良く問われる。現時点では脳の仕組みも完全に分かっていないし、強化学習がすぐに役立つわけでもない。しかし、脳が非常に強力なお手本であることは間違いないし、機能的な面からトップダウンで考える際には、工学的なセンスも必要であろう。両者の中間を進むと、研究の評価が難しく、中途半端になる危険も大きいが、大所高所からの研究ができるという意味で、両者のエッセンスを抽出し、**あえて中間を進む**ことが最も効率的であると考えている。そうすれば、さまざまな分野の考え方を導入することも容易である。そして、前述の知能と統合的学習の関係と同様、強化学習(知能)の研究には、各分野が分断されない統合的研究が重要と考える。

参考文献

[柴田 01-1] 柴田克成, 西野哲生, 岡部洋一: Actor-Q アーキテクチャに基づく能動認識学習システム, 信学論, Vol. J84-D-II , No.9, pp.2121-2130, 2001. 9.

[Shibata 02] K. Shibata and M. Sugisaka: Dynamics of a Recurrent Neural Network Acquired through the Learning of a Context-based Attention Task, Proc. of AROB 7th, pp. 152-155, 2002. 1.

[柴田 01-2] 柴田, 杉坂, 伊藤: 強化学習によるリーチング運動の獲得, 信学技報, NC2000-170, pp. 107-114, 2001. 3.

[前原 02] 前原伸一, 杉坂政典, 柴田克成: 移動物体の捕獲行動学習におけるセンサ動作の必要性, 信学技報, NC2001-153, pp.159-166, 2002. 3.

[柴田 99] 柴田克成, 伊藤宏司: 利害の衝突回避のための交渉コミュニケーションの学習と個性の発現, 計測自動制御学会論文誌, Vol.35, No.11, pp.1346-1354, 1999. 11.

[Shibata 00-1] K. Shibata, M. Ueda, and K. Ito: Emergence of Individuality and Sociality by Reinforcement Learning, Proc. of AROB 5th 2000, Vol. 2, pp. 589 - 592, 2000. 1.

[Shibata 00-2] K. Shibata and K. Ito: Autonomous Learning of Reward Distribution for Each Agent in Multi-Agent Reinforcement Learning, Intelligent Autonomous Systems, Vol. 6, pp. 495-502, 2000. 7.

[柴田 01-3] 柴田克成, 岡部洋一, 伊藤宏司: ニューラルネットワークを用いた Direct-Vision-Based 強化学習, 計測自動制御学会論文集, Vol.37, No.2, pp.168-177, 2001. 2.

[飯田 02] 飯田大, 杉坂政典, 柴田克成: Direct-Vision-Based 強化学習による実移動ロボットの行動学習, 信学技報, NC2001-154, pp. 167-174, 2002. 3.

[津本 86] 津本忠治: 脳と発達, 朝倉書店, 1986.

[畠中 92] 畠中寛: 神経成長因子ものがたり, 羊土社, 1992.