

可動カメラを用いたActor-Q学習による能動認識の学習

大分大学 ○AHMAD AFIF MOHD FAUDZI 柴田克成

Acquisition of active perception and recognition through Actor-Q learning using a movable camera

Ahmad Afif Mohd Faudzi and Katsunari Shibata, Oita University

Abstract: In the previous research, it was shown in simulation that active perception and recognition problem can be learned through Actor-Q learning. Appropriate camera motions and recognition timing as well as the correct recognition were acquired. In this paper, the authors tried to verify whether appropriate continuous camera motions and identification of two patterns could be acquired simultaneously through Actor-Q learning using a real camera. As a result, within a limited number of trials, these functions cannot be achieved. However, after three stages of learning were introduced, within the same number of trials, it was observed that active perception and recognition can be learned through Actor-Q learning.

1. はじめに

われわれはさまざまなセンサを用いて外界の情報を取り込むことによって、周りの状況を認識し、適切に活動している。しかし、外界の情報は膨大であり、すべての情報を細部まで獲得することは困難、かつ、効率が悪いと考えられる。人間は、目からの視覚によって最も多くの情報量を得ているが、能動的に頭や目を動かすことで、効率的に必要な情報を獲得している。また、網膜のセンサセルの分布は不均一であり、センサ全体を使って外界の様子を大まかにとらえ、中心部の分布が密な部分を適切な場所に移動して、効率的に認識を行っていると考えられる。

人間の目の動きと認識は非常に柔軟、かつ、知的な作業に見える。たとえば、本を読む時には、必ずしも文字を一文字一文字順番に切り出して認識するのではなく、読みながら文脈を理解し、次に出てくる言葉や内容を推測し、その状況に応じて目を動かすことで、効率的、高速な認識を実現しているように感じられる。このように、さまざまなことを同時並列に考慮し、柔軟な目の動きと認識の調和を実現できるのは、脳の並列性によるものと考えられる。

ビジュアルトラッキングの研究はすでに数多く行われているが、設計者がパターンマッチングやオプティカルフローに基づいて追跡する方法を予め与えることが一般的である。また、文字やパターン認識の研究も盛んに行われてきている。所定の領域に書かれた文字に対しては手書きの文字でも高い認識率が実現されている。しかしながら、柔軟な認識システムのために、認識の機能だけではなく、カメラモーションまで含めて総合的に学習させた研究は著者らが知る限りはない。

センサ動作と認識の両者を、調和を保ちながら同時

に学習させていく方法として、強化学習とニューラルネットワーク(NN)を組み合わせることが考えられる。強化学習は、報酬や罰による自律的な学習則であるが、一般的には行動やプランングのための学習ととらえられる傾向が強い。一方、ニューラルネットワークは非線形関数の近似器としてのメリットが注目されている。一方、筆者らのグループではニューラルネットワークに強化学習を適用することによって、センサからモータまでの、認識や記憶も含む一連の機能がシステム内で創発することを示してきた[1][2]。

この考え方に基づいて、視覚センサの不均一センサセルの密な部分を必要などところに移動させて、適切なパターンの認識を適切なタイミングで行うことができることがシミュレーションによって示されている[3]。またこの際、認識時の離散的意志決定と連続的なセンサ動作をともに強化学習によって学習するために、Actor-Q学習という学習方法が提案されている。

本研究では、実際の可動カメラを用いてActor-Q学習により能動認識の機能が獲得できるのかを検証した。[3]と同様に不均一な解像度の画像を用いてNNの入力とした。また、[3]と同様にフィードフォワードNNを用い、文脈を考慮した認識やセンサ動作の学習は今後の課題とした。

2. Actor-Q学習

Actor-Q学習は離散的な意志決定と連続的な動作を同時に学習することができる。この学習システムの出力は離散的な「行動」と、連続値出力である「動作」に区別される。システムは、始めに「行動」を決定し、その行動が「動作」を必要とする場合に、さらに「動作」を決定する。行動はQ値により決定し、Q-learning

で学習する。一方、「動作」の決定にはActor-criticのActorを用いる。そして、Actorの学習には、その行動の対応するQ値をCriticの代わりに用いる。

システムはFig. 1のようにQネットとActorネットの2つのニューラルネットで構成し、Qネットの各出力は各行動に対応させ、Actorネットの各出力はカメラの動作信号に対応させる。ここでは「パターン p であるという認識結果を出す」と「カメラを動かす」を行動として扱い、提示パターンが P 個ある場合は、個々のパターンであるという認識の結論を下す P 個の行動と、カメラ移動の1個をあわせた $(P+1)$ 個の行動があるとする。

選択された行動が「認識」の場合には、それが正解ならば報酬が、不正解の場合は罰が与えられ、その試行が終了する。そして、提示パターンが p であるという結論を下す行動を“*Recog:p*”と書くと、その行動のQ値に対する教師信号は、

$$Q^{“Recog:p”,train} = r \quad (1)$$

と与えられる。ただし、 r :報酬であり、正解であれば0.9とし、不正解であれば0.1とした。一方、「カメラ動作」が選択された場合は、Actorネットの出力によってカメラを移動させる。この場合は、試行は終了せず、カメラ移動後に新しい入力を得て、再び次の行動が選択される。「カメラ動作」の場合の教師信号は、

$$Q^{“motion”,train} = \gamma \max_{a \in A} Q_a(s_{t+1}) \quad (2)$$

で与えられる。ここで、 γ :割引率、 A :選択可能な行動選択の集合である。

カメラの移動 \mathbf{m} はベクトルであり、カメラのPan(横方向の移動)、Tilt(縦方向の移動)の移動量を規定するものとする。そして、Actorネットの出力 \mathbf{O}_m に乱数 \mathbf{rnd} を加え、定数 K を掛けた

$$\mathbf{m} = K(\mathbf{O}_m + \mathbf{rnd}) \quad (3)$$

にしたがってカメラを動作させる。Actorネットの出力 \mathbf{O}_m の教師信号は

$$\mathbf{O}_{m,train} = \mathbf{O}_m + (\gamma \max_a Q_a(s_{t+1}) - Q^{“motion”,train}(s_t)) \mathbf{rnd}. \quad (4)$$

とした。QネットとActorネットは上の教師信号を用いて誤差逆伝搬 (Error Back Propagation, BP) で学習させる。

3. 問題設定と学習

本論文では、視覚センサとして可動カメラCanon VC-C50iを用いる。数字‘0’と‘9’の2つのパターンを用意し、モニタ上に提示し、どちらのパターンであるかを識別する問題を扱う。また、Fig. 2のように、中心部の解像度が大きく、周辺部の解像度が小さいという

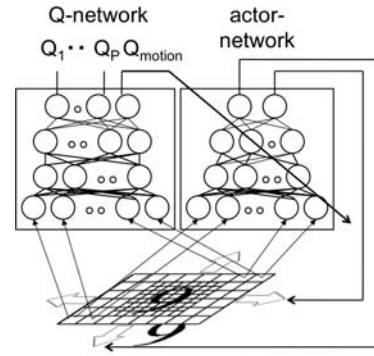


Fig. 1 Actor-Q based active perception and recognition learning system

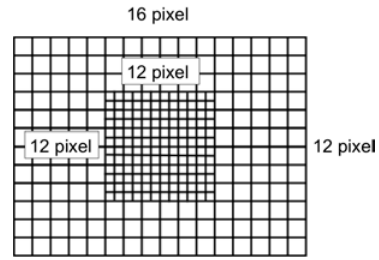


Fig. 2 Visual image with non-uniform resolution

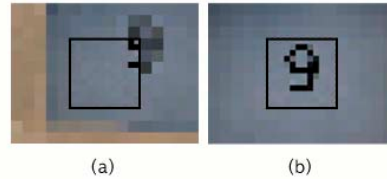


Fig. 3 The difference of the captured image depending on the part that captures a presented pattern.

不均一性を導入した。学習中に、提示されるパターンとカメラの初期位置は毎試行ランダムに設定するが、それらの情報は与えない。Fig. 3(a)のように、パターン全体が視覚センサの中に入っているにもかかわらず、粗い部分で捕らえた場合は認識することが難しいため、より良い認識ができる高い解像度の領域で捕らえるように、カメラを移動させてから、認識の結論を下すことができる。そして、その認識結果が正しいかどうか強化信号として与えられる。そして、強化学習することによって適切なカメラ移動、認識と認識結果を出力するタイミングを獲得する。

次に、システム設定について述べる。Fig. 1のように、システムは2つのネットを持ち、それぞれを4層とした。各層のニューロン数は、入力から、Qネットの方が900-250-20-3、Actorネットの方が900-250-20-3とした。2つのネットはともに、全体の16×12の粗い画像の中央部を12×12の解像度の高い画像で置き換えた全部で300ピクセル(RGB)の画像の信号を-0.5から0.5に線形変換した値を入力とした。

Qネットには3つの出力があり、それぞれ、「0と認識

結果を出力する”、“9と認識結果を出力する”と“カメラ動作”の3つの行動のQ値に対応している。一方、Actor ネットは、カメラ動作を決めるPanとTiltの2つの出力を持っている。各ニューロンの出力関数は-0.5から0.5を領域とした。また、Q値とニューラルネットの出力については、Eq. (1)に述べた0.1から0.9の範囲のQ値が-0.4から0.4のニューラルネットの出力になるように、お互いに線形変換して用いた。したがって、実際の教師信号は、Eq. (1)で求めた値から0.5を引く必要がある。また、教師信号は-0.4から0.4の範囲を超えないようにした。

中間層から出力層の初期重みは0.0とし、それ以外は-0.1から0.1の乱数とした。Eq. (2)と(4)のQ値の学習に用いる割引率 γ は0.986、Eq. (3)のカメラの最大移動量は 2.25° とした。また、試行錯誤のために加えた乱数 \mathbf{rnd} は1つの一様乱数を3乗したものとし、その値域は ± 1.0 からスタートし、学習につれて徐々に小さくして行った。行動選択には、学習中はQ値が最大の行動を選択する ϵ -greedy選択を用い、 ϵ の値はEq. (5)にしたがって徐々に下げていった。

$$\epsilon = \exp(-\text{trial} \times 0.00017) \quad (5)$$

ただし、 trial は試行回数である。

4. 学習結果

10,000試行の学習を行なったが、正しく認識できるようにはならなかった。Fig. 4に、(a) パターン‘0’、(b) パターン‘9’を提示した場合のそれぞれについて、試行が終わったときのQ値の変化を示す。学習が成功すれば、たとえば、パターン‘0’が提示されたときは、‘0’と結論を下すQ値(赤の+)が大きくなり、‘9’の結論を下すQ値(緑のx)は小さくなる。Fig. 4を見ると、Q値の変化があまりないことがわかる。

先行研究[3]では、学習に100,000試行要しており、学習回数を増やせば学習できる可能性はあると考えられる。しかし、実機を用いているので、学習にかかる時間やモータへの負荷などから、数万回をやるのは困難である。そこで、カメラ動作とパターン認識を個別に学習させた後に、組み合わせた学習を行なうという3段階の学習を行なった。

最初の段階では、カメラ動作を学習させた。システムがカメラを移動することによって、提示されたパターンを画像の中心に持ってくると報酬がもらえるタスクである。カメラの初期位置はランダムに設定した。レベルが上がるほど、初期位置の領域が遠くなり、最短で報酬がもらえる状態に到達するステップも増える。ここでは、Actorネットとカメラ動作に対応するQ値のみを学習させ、actor-criticのアルゴリズムを用いた。

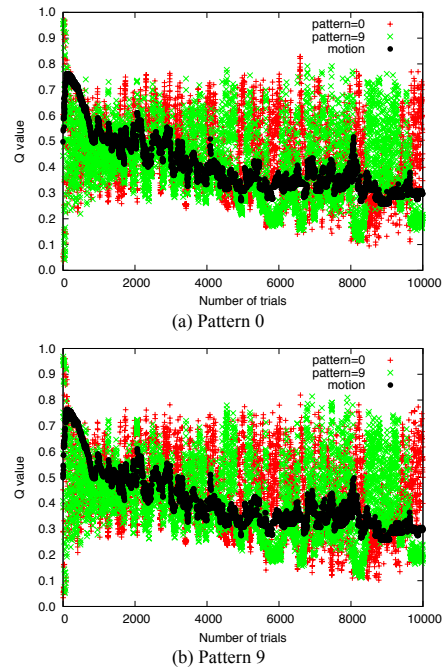


Fig. 4 The difference of the captured image depending on the part that captures a presented pattern.

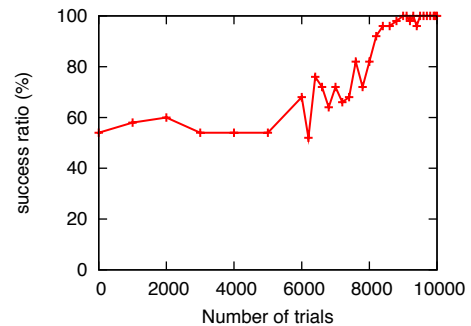


Fig. 5 Learning curve for active perception and recognition task after two preliminary learnings.

次に、パターン認識のタスクを学習させた。ここでは、パターン‘0’と‘9’がモニタ上に提示され、正しくパターンを区別すると報酬がもらえるタスクである。最初の2000試行は、カメラの初期位置を固定とし、パターンが画像の中心に捕らえている状態で、その後、初期位置を中心から ± 2.5 領域でランダムとした。このタスクでは、Q-learningを用いて、パターン認識に対応するQ値のみを学習させる。

最後の段階では、両方のネットにカメラ動作と認識両者を同時に学習させた。カメラの初期位置をランダムとし、ある試行回数達したら、その領域を大きくした。Fig. 5に、学習中に保存した重み値を用い、行動決定における乱数要素をなくし、50回試行させたときの試行回数に対する成功率の変化を示す。学習が進むに連れて成功率が100%に近づいていることがわかる。

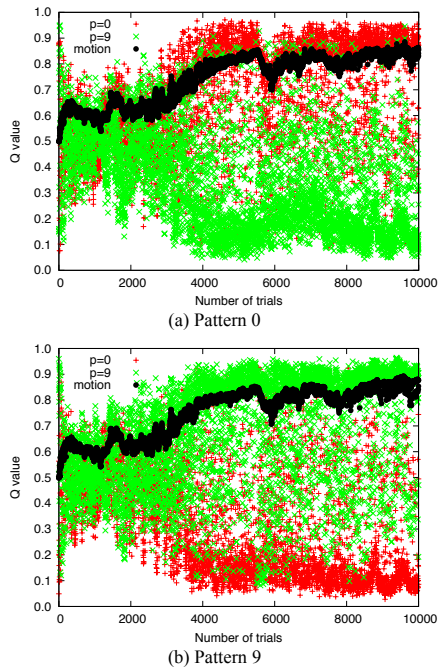


Fig. 6 The change of Q-values at the end of every trial in the learning of active perception and recognition task after two preliminary learnings.

Fig. 6に、毎試行の終わりの各行動に対応するQ値の変化を示す。パターン‘0’が提示されるときに、(a)では、パターン‘0’に対応するQ値を表す赤点は他のQ値より大きくなり、パターン‘9’に対応するQ値を表す緑点は小さくなっていった。この差より、システムは正確に認識できた。しかし、学習が終わったところでも、間違った認識をしている点が少し見えるが、これは行動選択に用いる ϵ -greedyの影響で、最大Q値でない行動を選択したためと考えられる。

Fig. 7に、学習終了後に乱数を切って行動させたとき、初期位置から認識結果の結論を下すまでのカメラの動作例を示す。ステップ4では、パターンが既に画像の解像度の高い部分に捕らえられているが、その後、より良い認識ができる位置を探すため細かい動きをしている。そして、ステップ9で正しい認識結果を出力した。

Fig. 8(a)はFig. 7と同じ試行のQ値の変化を表す。図を見ると、カメラ動作に対応するQ値は最初から大きいことがわかる。それで、次第にパターン‘9’対応するQ値との差が小さくなって、ステップ9で認識のQ値の方が大きくなり認識する。ステップ4からすでに9のパターンを認識するQ値の方が0のパターンを認識するQ値よりはるかに大きくなっているのでもう少し学習させればより早く認識できるようになるのではないかと考えられる。また、(b)に、カメラを移動するためのActorの出力値の変化を示す。Panの値が正で、Tiltの値が負であることから、Fig. 7のようにシステムがカメラを

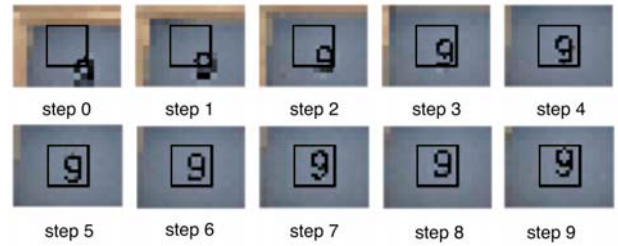


Fig. 7 A series of images from an initial position to the goal where the recognition result is decided after learning.

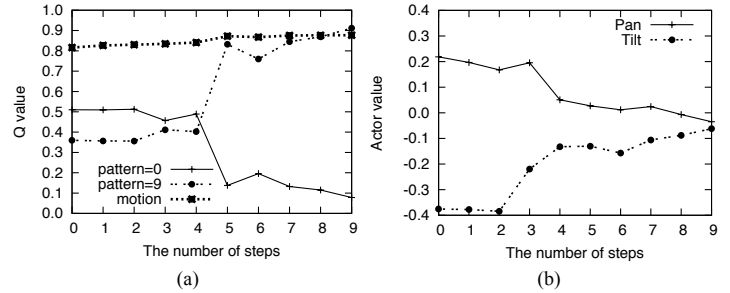


Fig. 8 The change of Q-values and actor outputs in the trial as shown in Fig. 7

右下に動かしていることがわかる。

5. まとめ

本研究では、可動カメラを用いて、Actor-Q学習により、*pan*, *tilt*の二次元のカメラ動作をしながら0と9のパターンを識別する能動認識の学習を行なった。システムをゼロから学習させたところ、10,000試行の学習を行なったが、正しく認識できるようにはならなかった。そこで、最初にカメラ動作と認識を別々に強化学習で学習し、その後両者を同時に学習した。その結果、学習後、システムはカメラを適切な場所に移動し、正確に認識することができた。

今後の課題として、パターン数を増やし、照明条件の違う昼間と夜間中に学習することが挙げられる。さらに、文脈を利用したセンサ動作と認識を獲得するため、リカレントニューラルネットワークを組み込むことが考えられる。

参考文献

- [1] K. Shibata and T. Kawano, Acquisition of Flexible Image Recognition by Coupling of Reinforcement Learning and a Neural Network, SICE JCMSI, Vol. 2, No. 2, pp. 122-129, 2009.
- [2] H. Utsunomiya and K. Shibata, Contextual Behavior and Internal Representations Acquired by Reinforcement Learning with a RNN in a Continuous State and Action Space Task, Advances in Neuro-Information Processing, Lecture Notes in Computer Science, Vol. 5507, pp. 970-978, 5507-0970.pdf (CD-ROM), 2009.
- [3] 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, 電子情報通信学会論文誌, Vol. J84-D-II, No. 9, pp.2121-2130, 2001.