

画像入力ニューラルネットを用いた強化学習による 可動カメラの対象物検出行動の獲得

大分大学 河野友彦 幸和芳 柴田克成

Acquisition of Object Detection Behavior of a Movable Camera by Reinforcement Learning using Neural Network with Image Input

Tomohiko Kawano, Kazuyoshi Yuki, Katsunari Shibata, Oita University

Abstract: In this paper, the authors point out the possibility that important information is lost by the feature extraction or some other image processing. In order to solve the problem, the authors advocate that the visual signals are put into a neural network directly as inputs and the neural network is trained according to some learning algorithm. In order to examine the utility of the approach, an experiment using two AIBO robots was done. One AIBO caught an image including the other fixed AIBO with a variety of light conditions and backgrounds, and decided its action according to the outputs of the neural network whose input is raw visual signals. It could learn to face the front of the other AIBO and bark after some appropriate head rotations with 90% success rate.

1. まえがき

人間は実環境において、視覚情報から複雑な形や色の対象物を容易に検出することができるが、ロボットに同様のことをさせようとすると、照明や背景の変化も考慮しなければならないなど、難しいタスクとなる。画像から対象物を検出する際には、画像処理によって、明るさを補正し、特徴抽出を行ない、抽出された特徴から対象物を検出したり、さまざまな見え方のテンプレートとのマッチングを行なうのが一般的である。しかし、これらの手法が本当に最も適切な手法であると言えるのであろうか？

対象物の検出や画像認識の過程は、画像信号を入力とし、検出、認識結果を出力とする関数であると考えることができる。そして、この関数を目的に沿っていかに最適に近いものにするかが問題である。人間は意識しないところで画像の明るさや色の補正をする機能を有していることから推測すると、実は、意識に登らないところでさまざまな処理を並列に行っているからこそ、実環境でも容易に対象物の検出や画像認識を行うことができるのかもしれない。一方、前述のロボットに与える特徴抽出では、非常に膨大な画像の情報の中から、人間が必要と考える特徴を抽出し、そして、要らないと思われる情報を捨てている。しかし、一見不要ではないかと思われる情報の中にも有効な情報が含まれており、実はそれを有効に利用できるかどうかはわれわれ人間の柔軟な認識能力とロボットの認識能力との大きなギャップになっている可能性があるのではないかとわれわれは考えている。

では、どうやって膨大なセンサ情報を使いこなせば良いかという問題になる。前述のように、画像認識は画像信号を

入力する関数の最適化問題であると考えられる。そうなれば、画像信号をニューラルネットに入力し、何らかの学習をさせれば、そのための最適化が行なわれ、結果的に、人間のよう柔軟な検出・認識能力を実現できるのではないかと期待される。

画像信号を入力として強化学習を行う際には、従来は前処理をして状態分けを行った後に強化学習を適用する方法が一般的であった。これに対し、われわれの研究室では、Fig.1のように、画像を始めとするセンサ信号をニューラルネットに直接入力し、強化学習で学習させることで、必要となるさまざまな機能がニューラルネット内に創発し、それが知能形成につながると主張してきた[1]。そして、実ロボットによる箱押し行動を画像処理等の知識を一切与えずに画像から適切な行動を学習できることを示した[2]。しかし、周りは白い紙で囲うという特殊な環境下で、さらにカラーではなくグレースケールの画像を入力としていた。

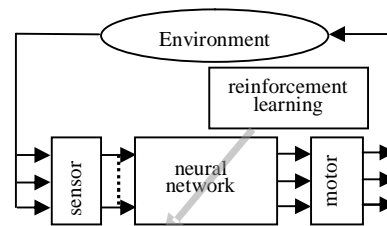


Fig.1 Reinforcement learning using a neural network with direct sensory inputs.

本論文では、当研究室で用いてきた上記の手法が、強化学習による対象物検出や画像認識過程の最適化につながることで、従来実現の難しかった人間のような柔軟で優れた認識機能を獲得できる可能性があるのではないかとこの点に焦点をあてる。そして、照明条件や背景が異なるさまざまな条件下で、従来より複雑な対象物の認識を学習させることができるかどうかを確認する。また、ニューラルネットの構造による学習能力の違いについてもあわせて確認を行った。

2. タスク設定

本論文では、2台のSONY製犬型ロボットAIBOを用いてタスクを行わせた。両者はFig.2のようにステーションに置いて43cm離して向かい合わせに固定し、そのうち1台(黒)は首を振ることで可動カメラとして用い、もう1台(白)は少し複雑な形をした動かない対象物として用いた。カメラは35万画素で、水平画角は約54°である。カメラの原画像は208×160のRGBデータであるが、AIBOから送信する前にXY方向で4ピクセルごとに間引きを行い52×40と16分の1のサイズに縮小したものを学習に用いた。カメラの水平可動範囲は、正面を向いている時を0°としたとき±89.6°であるが、カメラ画像の範囲内に対象物が収まるように、±20°の範囲を5°の間隔でスイングさせる。そのため、状態数は9状態となる。

可動カメラとして用いる黒のAIBOは、画像をニューラルネットに入力した際の出力に従って行動を行う。行う行動は、“首を右に動かす”“首を左に動かす”“吠える”の3行動とし、首を動かす場合は5°該当の方向へ回転させた。初期位置は、1試行ごとにランダムで、その位置から首振りを開始する。そして、もう1台のAIBOが正面に見えたとき(状態0)に吠えると報酬がもらえ、それ以外のときに吠えると罰を受ける。たとえば、Fig.3のように正面より右に10°の方向を向いている状態2から開始する場合は、首を左に2回転かして、正面を向いてから吠えるという行動をとると報酬をもらうことができる。



Fig.2 The experimental environment.

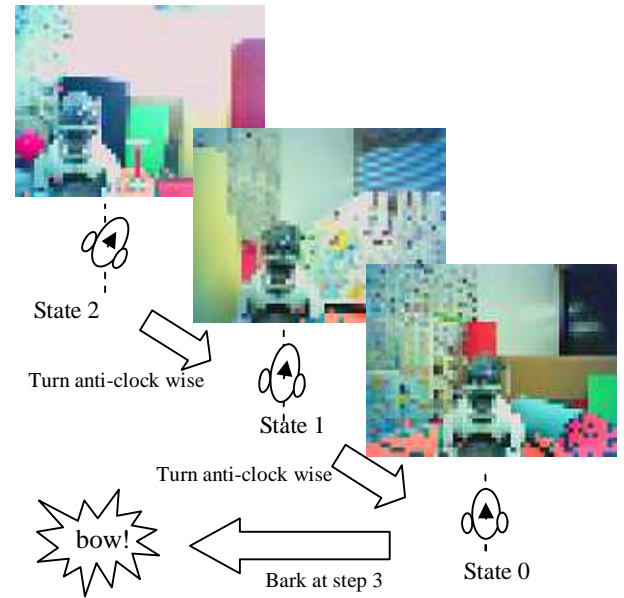


Fig.3 The robot behavior required in the task. Each of the images on the upper of picture is a sample image for each corresponding state.

3. 学習方法

本論文では、実ロボットに搭載されたカメラ画像を入力とするニューラルネットワークを強化学習によって学習させ、対象物検出行動の獲得を行う。本当は、実機上で学習を行いたいところであるが、学習に非常に時間がかかるため、やむをえず学習用にさまざまな画像をあらかじめ採取しておき、実機を用いずに学習を行った。

画像の採取時は、外が明るい昼間、夕方、暗くなった夜、ブラインドの開閉などの照明条件を変え、さらにロボットの後ろにさまざまな物や壁紙を適当に配置した。そして、9状態それぞれに1枚ずつ画像を採取して1パターンとし、条件を変えることによって同じ状態でもパターンごとに明るさ、背景が異なるようにした。

学習用サンプルとして312パターン2808サンプル、パフォーマンスのテスト用に92パターン828サンプルをシミュレーションに用いた。Fig.3に実際に入力される画像の例を示す。各ステップごとに背景が変わっていることがわかる。また、この画像は上から、昼間、夕方、夜の各時間帯に採取した画像で、各画像右上部に窓が写っており、時間が変わると見え方が変わってくるのがわかる。

ニューラルネットワークの入力は、52×40の2080Pixelに、RGBの3bitを掛けた計6240個の信号とした。RGB256階調を0.0~1.0に正規化した後に、0.1を反転した値をニューラルネットワークに入力する。反転させるのは、背景が壁(白)の場合に入力信号の大部分が1に近い値となり、学習が不安定になることを防ぐためである。

シミュレーションでは、画像はステップごとに312パター

ンの中からランダムに選択した。ニューラルネットワークの出力は、3 つ用意し、“吠える” “右へ首を振る” “左へ首を振る” の各行動の Q 値として扱った。そして、ε-greedy で実際に行う行動を選択した。探索率は 0.13 で固定とした。ニューラルネットワークは、階層型で層数を 3~5 層で変化させ、層数による検出成功回数の変化を比較した。それぞれのユニット数は、最下層から 6240-1000-3(結合数は約 624 万個)、6240-800-150-3(約 511 万個)、6240-600-150-40-3(約 384 万個)とし、多層になるほど学習時のパラメータ数を小さくした。出力関数は-0.5~0.5 のシグモイド関数を使用し、学習係数は 0.5 とした。

強化学習のアルゴリズムは Q-Learning を使用した。ニューロンの出力関数として用いたシグモイド関数の値域は -0.5 ~ 0.5 としたので、その値を -0.1~0.9 に線形変換して、Q 値として用いた。これは、Q 値は 0 以上となることを想定しているため、出力の値域を有効に利用するためである。割引率は 0.8 とし、正面を向いたときに吠えらるともらえる報酬は 0.8 とした。

ニューラルネットワークの学習は、行動後の画像を入力したときの出力の中から最大の Q 値を求め、1 ステップ前の時間にさかのぼった後に、行った行動に該当するニューロンの出力に対して以下のような教師信号 *TeachSig* を与えて行う。通常の状態遷移では

$$TeachSig = \gamma \times (\max_p Q(S_{t+1}, p) + 0.4) - 0.4$$

状態遷移の結果報酬 *r* が得られたときには

$$TeachSig = r - 0.4$$

行った行動が失敗であったときには

$$TeachSig = Q(S_t, a_t) - 0.02$$

(γ : 割引率、 a_t : 時刻 t でとった行動、 S_t : 時刻 t での状態、 $Q(S_t, a_t)$: 時刻 t で状態 S_t で取った行動 a_t に対しての Q 値)

という式で教師信号を与え、通常の誤差逆伝搬(BP)法で学習を行わせる。このとき、該当しない行動の出力は、学習させなかった。また、教師信号は-0.4 より小さくならないとした。

学習を 100 試行行うごとに、パフォーマンスのテストとして $\epsilon=0$ (ランダム成分を 0) で学習をしない状態で 50 試行探索を行った。このテストでは、学習に使用したものと同じパターンと、汎化能力の指標とするために、学習に全く使用していないパターンで別々にテストを行った。テストでの初期位置は、学習時のランダム初期位置とは異なり、試行ごとにゴールから一番遠い状態 -4 と状態 4 に交互に切り替えるようにした。

4 . 実験結果

最初にシミュレーションによる実験を行った結果を示す。Fig.5 は学習パターン、Fig.6 はテストパターンを用いて、パフォーマンステストを 3, 4, 5 層のそれぞれの場合で行った結果を示す。グラフの縦軸は、学習中 100 試行ごとにパフォーマンステストを 50 試行を行い、1000 試行分の成功回数の平均をとったもので、横軸を試行回数とした。なお、実機では最短ステップでゴールしなくても成功すれば成功回数にカウントするように設定しているため、ここでは 13 ステップ以内に間違っずに吠えずにゴールすれば成功とした。13 ステップという数字は、たとえ状態 4 から試行を始めて間違っずに状態 -4 まで行ったとしてもゴールへたどり着くことのできるステップ数である。

テストの結果、学習が進むにつれ、学習パターン、テストパターン双方とも検出率は 9 割程度に収束しており、学習に用いているパターンでも期待された 10 割近くには届かなかった。一方、未知のパターンであるはずのテストパターンを用いてのテストでは、学習パターンの場合に比べ若干低い程度で極端な差はなかった。このことからニューラルネットワークの汎化能力がある程度有効に働いており、未知の画像にも対応できていることがわかる。

次に、ニューラルネットワークの構造による学習状況の違いを比較する。一般的に、層数が多くなるほどニューラルネットワークの汎化能力は上がりやすく、未知の画像を入力しても検出ができる可能性が高くなることが期待される。そのため、層数の変化による汎化能力の向上によって検出率に影響が出るかを見た。

Fig.6 のテストパターンでのパフォーマンステストで、3 層の場合のみ検出率の伸びが悪く、学習が遅くなっているが、これは Fig.5 の学習パターンの場合と同様な傾向である。したがって、これは汎化能力の問題ではなく 3 層では学習自体が遅かったと考えられる。これは、多層化すると学習が遅くなるという通常と逆の傾向であり、今後検討が必要である。一方、層数の増加による汎化能力の向上は特に観察されなかった。これは前述のように、このタスクでは 3 層でも十分に汎化能力の効果が得られたためではないかと考えられる。

次に、シミュレータによる学習時に得られた重み値を使用し、100 試行だけ実機で動かしてみた。シミュレータでは 1 ステップごとにパターンをランダムに変化させていたが、実機では 20 試行ごとに背景を変えた。つまり、100 試行行うので 5 パターンの背景を与えたことになる。また実験中は、20 試行ごとにブラインドの開け閉めも行い照明条件を変えながら行った。また、外がまだ明るい昼間と真っ暗になってからの夜とで 2 回実験を行った。

実験には、サンプル画像を採取するときに使用したときと同じ AIBO を用いた。AIBO は、無線 LAN 経由でデータ

のやりとりができ、AIBO 側からは画像データを、PC 側からは首の角度指示値をそれぞれ送受信できるようにした。AIBO 側は、指示値に従った首の制御、画像送信のみを行い、PC 側では送られてきた画像信号をニューラルネットワークに入力し計算処理を行い、行動選択をする。その行動に従った角度指示値を PC 側から AIBO 側に返す。

使用した重み値は、先ほどの層数による比較を行った際に 5 層のネットワークで 2 万試行学習したものをを使用した。試行後の結果を Table.1 に示す。表中の 3 つの数字は、成功した回数、探索率 の影響で吠えるべきところではないところで吠えてしまった回数、ニューラルネットワークの出力が最大となる行動を取った結果、間違えて吠えてしまった回数を表す。

Table.1 を見ると、実機で学習をしなかった場合、昼間でも夜でも約 9 割の検出率となっており、シミュレータでの学習によって実機でもある程度は検出行動を行うことができた。しかし、学習させながら試行させた場合には、探索率による試行錯誤によって失敗した回数を除いても成功率は 8 割程度に留まっており、によるランダム成分が入ることで、元々の学習が少し崩れてしまったのではないかと思われる。これについては、実機上でもう少し多くの学習をさせてみる必要がある。

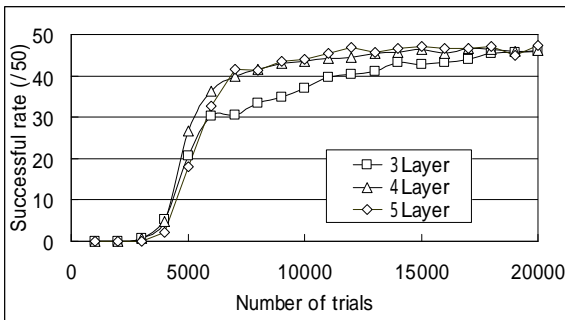


Fig.5 Comparison of learning curve according to the number of layers. (Learning Pattern)

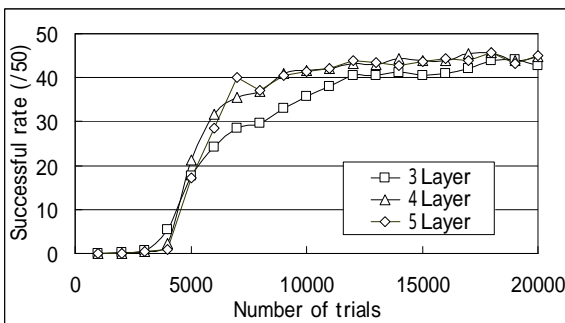


Fig.6 Comparison of learning curve according to the number of layers. (Test Pattern)

Table 1 Success rate in 100 trials using the real robot.

Successive three numbers indicate [success]-[failure (due to exploration factor)]-[failure].

	No learning (0.0)	with learning (0.13)
Daytime	95-0-5	69-17-14
Night	90-0-10	64-21-15

5. あとがき

本論文では、照明条件や背景が変化する実環境に近い実験環境において、AIBO という少し複雑な形状の物体の検出タスクを扱った。そして、6000 個強のカラー画像信号を入力するニューラルネットワークを用い、AIBO を正面に捕らえたときに吠えると報酬がもらえるという設定で学習を行った。そして、実機において背景や照明条件を変えても 90% 程度の成功率を得ることができた。未知のパターンでもある程度対応できたことから、画像信号を直接ニューラルネットワークへ入力し学習することで人間のような柔軟な認識機能を獲得できる可能性を示唆することができたと考えている。

しかしながら、本論文で扱ったのは、AIBO 同士を固定した状態で首だけを動かして認識を行わせるというまだまだ簡単なタスクである。今後は、AIBO 自身が動き回り、背景や対象物の見える方向が変化する状況のように、人間が認識方法を与えることが明らかに困難と考えられる問題でも、提案している学習法を使用することで柔軟な認識を行えるかどうかについて検証していきたい。

謝辞

本研究は、日本学術振興会科技学術研究費補助金基盤研究 #15300064 and #19300070 の補助を受けた。ここに謝意を示す。

6. 参考文献

[1] 柴田克成, 飯田大, "視覚センサ付き実ロボットによる箱押し行動の獲得 - 強化学習によるセンサ - モータ間トータル機能獲得への第一歩 -", 第 14 回インテリジェント・システム・シンポジウム (FAN Symposium) 講演論文集, pp. 123-128, 2004.10

[2] 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 - センサからモータまで -", 計測自動制御学会論文集, Vol.37, No.2, pp.168-177, 2001.2