物体移動タスクにおける繰り返し予測に基づく行動生成

Action Production Based on Repetitive Prediction in an Object Pushing Task

板舛尚樹[*1] 本多透[*2] 岡夏樹[*2]
Naoki Itamasu Toru Honda Natsuki Oka

[*1] NEC ネッツエスアイ株式会社
NEC Networks & System Integration Corporation

[*2] 京都工芸繊維大学
Kyoto Institute of Technology

In the real world, observed information fluctuates continuously for various reasons even in a steady environment, and it is necessary for them to being robust for the fluctuations. Humans seem to cope with this problem by predicting the future state based on the past experience, and deciding actions based on the prediction. In addition, humans often decide actions based on not only the near future but also the far future.In this study, we propose a method for predicting the far future through repetitive prediction of the near future, and deciding actions based on the prediction. The effectiveness of the proposed method was demonstrated in a experiment in which a small robot e-puck pushes an object to a goal. We also clarify the condition in which the prediction is necessary by experiments in several tasks of different difficulties.

1. はじめに

自律ロボットには，自身のセンサより得られる外部環境の観測情報から，適切な出力を生成する行動生成モデルが必要とされる．行動生成モデルの設計は，環境やロボットの動特性が複雑になるに従い困難となるため，近年，学習により自動的に行動生成モデルを獲得する手法が研究されている．行動生成モデルを構築する上で，多様な環境への速やかな適応の実現が重要視されている．ここで，人間は様々な過去の経験を利用することで，未来の状態を予測し，その予測状態と現在の状態に基づいて行動決定をしていると考えられる．また，人間は近い未来だけではなく，遠い未来まで予測して現在取るべき行動を決定することが多々ある．実際に現在までに予測情報を行動生成に利用するモデルが多数提案されてきた[Tani 1999] [Ohigashi 2003]．本研究では近い未来の予測を繰り返すことによって遠い未来の予測を行い，その予測情報を行動生成に利用する手法を提案し，小型ロボット e-puck を使用した実機実験に適用した．

2. 提案手法

本研究の提案手法では，予測器が現在の状態から 1 ステップ後の予測を行い，その状態を用いて更に予測を繰り返すことによって遠い未来の予測を実現する．そして，予測された状態が望ましい状態であれば，その予測情報に基づいて現在取るべき行動を決定する．以後，時刻 t の状態を S(t)と定義する．以下に提案手法における行動決定までの流れを大まかに説明する(図 1)．

この手法は行動候補生成器により現状態での行動候補を生成し、ひとつに絞りきれない場合、生成された行動候補ごとに次状態の予測を行う．予測器により出力された状態を状態評価器により評価し、ゴールでない場合、同様の手順を出力された各状態に適用し、ゴール状態にたどりつくまで予測を繰り返す．ゴール状態まで予測ができたら、その状態から現状態まで逆にたどっていき、ゴール状態へとたどり着く行動候補を行動として出力するという手法である．

2.1 行動候補生成器について

行動候補生成器には状態 S(t)が入力され，この状態の時に取るべき行動の候補 $a_1(t)$, $a_2(t)$, …を出力する．出力層に閾値を定めて，発火した出力ニューロンの数だけ行動候補を出力する．現

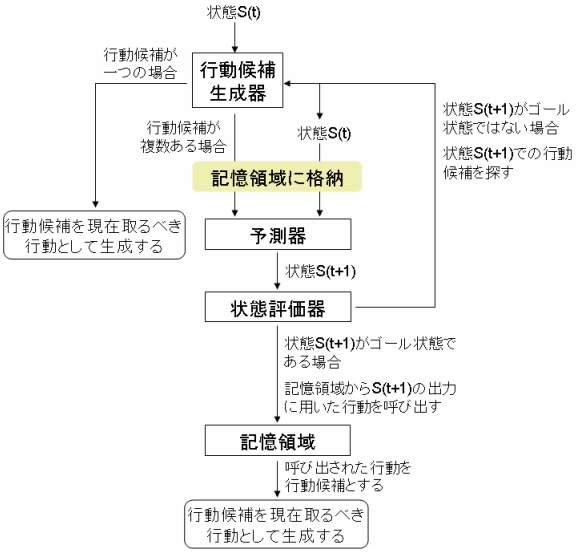

図 1 提案手法における行動決定までの流れ

在の状態 S(0)が入力されて，出力された行動候補が一つである場合，その行動を現在取るべき行動として決定する．状態 S(0)が入力されて複数の行動候補が出力された場合や，繰り返し予測を行っている途中の場合は，状態 S(t)と行動候補 $a_1(t)$, $a_2(t)$, …を記憶領域に格納する．学習には，ロボットが与えられた環境下で合目的的な動作を実現した際に得られる入出力(センサ-モータ)時系列データを用いる．

2.2 予測器について

予測器は記憶領域から受け取った状態 S(t)と行動 a(t)から次状態 S(t+1)を出力する．これは S(t)という状態において a(t)という行動を取ると次にどのような状態になるかを予測するものである．行動候補生成器と同様に，学習には入出力時系列データを用いる．

2.3 状態評価器について

状態評価器では予測器が出力した次状態 S(t+1)をゴールであるかどうか判断する．タスクを達成した際の状態 $S(t_{end})$を事前に知っておく必要があり，$S(t_{end})$と予測器が出力した次状態を比

連絡先：本多透，京都工芸繊維大学 大学院工芸科学研究科，
e-mail：m9622032@edu.kit.ac.jp

較することによってゴールであるかどうかの判定を行う．予測された次状態がゴールであると判断した場合，記憶領域に行動生成を命令する．

2.4 記憶領域について

記憶領域では状態 S(t)と行動候補生成器から出力された行動候補 $a_1(t)$, $a_2(t)$, …を図 2 のように，状態をノード，行動をリンクとして木構造の形で記憶しておく．幅優先探索を行い，次に探索すべきノードの親であるノードが示す状態と，それらを結ぶリンクが示す行動を予測器に渡す．状態評価器から行動生成の命令があった場合は，ゴール状態を示すノードに向かって根から伸びているリンクが示す行動を現在取るべき行動として行動生成を行う．図 2 のようにゴール状態が発見された場合，行動 $a_2(0)$を現在取るべき行動として生成する．

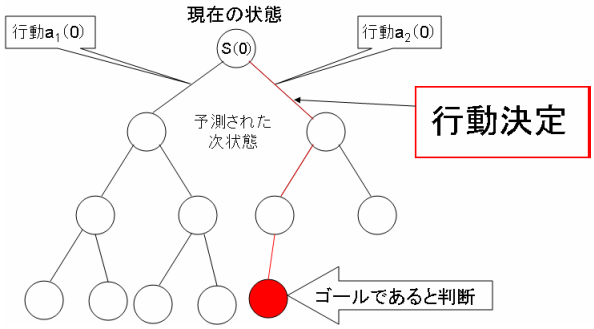

図 2 記憶領域における木構造

3. 実験方法

本研究では小型ロボット e-puck が直方体の物体をゴールである光源まで押して運ぶというタスクで実機実験を行った．e-puck は円筒形のロボットであり，直径 70mm，高さ 50mm となっている．e-puck ロボットの周辺に 8 個のセンサが配置されている．その配列は均等ではなく，ロボットの前面にはより多いセンサが配置されている(図 3)．

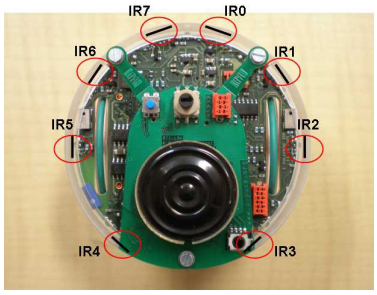

図 3 e-puck におけるセンサの位置

各センサは 3～4 センチ有効の赤外線近接センサおよび光センサの機能を持っており，物体までの距離と光の強さを感知することができる．実験環境の横幅は 600mm，縦幅は 420mm であり，光源は実験環境より 45mm 高い位置に設置した．直方体の物体は紙で作られており，その幅は 130mm，高さと奥行きは共に 32mm である．直方体の物体は中心から少しでもずれた位置を押すと，物体が左右に傾くことになる．e-puck はその都度位置を修正しながら，ゴールまで物体を押して運ぶことが求められる．

3.1 状態認識について

e-puck に搭載されているセンサを用いて，物体および光源までの距離を取得する．今回の実験では，赤外線近接センサ(IR0～IR2 および IR5～IR7 の 6 種類)の値と光センサ(IR0～IR2 および IR5～IR7 の 6 種類)の値を用いて，12 次元の情報から状態認識を行う．また，学習器への入力にはセンサの種類ごとに平均が 0，分散が 1 になるように正規化した値を用いた．

3.2 行動について

今回の実験で設計した行動は以下の 4 種類である．
前進：e-puck を 2cm 前進させる．
後退：e-puck を 2cm 後退させる．
左回転：e-puck をその場で左方向に 30 ° 回転させる．
右回転：e-puck をその場で右方向に 30 ° 回転させる．

3.3 行動学習について

e-puck を操縦して物体を押しながらゴールまで運ぶことにより，環境情報と行動がセットになった時系列データが得られる．この時系列データを複数試行分用意し，それらを繰り返し学習器に入力することによって学習を行う．各行動の教師信号は以下のように値を設定した．
前進：(0, 0, 0, 1)
後退：(0, 1, 0, 0)
左回転：(1, 0, 0, 0)
右回転：(0, 0, 1, 0)
つまり，ある入力が行われた時に前進させたとすると，この入力に対して(0, 0, 0, 1)が教師信号として与えられ学習を行う．

4. 実験タスクについて

本研究では予測が必要な環境を明らかにするために，以下のように難易度の異なるタスクを設定し，段階的に実験を行った．
[タスク 1] 単体ニューラルネットワーク(以下 NN)で学習が可能なタスク
[タスク 2] 単体 NN では学習が難しいがモジュール型ネットワークを用いることによって比較的容易に学習が可能なタスク
[タスク 3] パラメータの設定によってはモジュール型ネットワークを用いても学習が難しいが，予測を行うことによって適切な行動生成が可能なタスク
今回の実験では，代表的なモジュール型ネットワークである Mixture of Experts(以下 ME) [Jacobs 1991]を用いた．また，NN は全て Back Propagation(BP)法を用いて学習を行った．

4.1 タスク 1

タスク 1 では図 4 のように e-puck と物体の初期位置を設定した．最初から e-puck の目前に物体があり，その延長線上にゴールがある環境である．初期位置から e-puck を操縦して物体を押しながらゴールへ運ぶまでを 1 試行として，5 試行分の時系列学習データを準備した．5 試行中に 54 回分の入出力データが得られた．5 試行分の学習データを入力し学習を行うことを学習回数 1 回とし，単体 NN を用いて 2000 回の学習を行った．入力層のノード数 12，中間層のノード数 20，出力層のノード数 4 の 3 層 NN を使用し，学習率は 0.1 とした．学習後，初期位置に e-puck と物体を置き，単体 NN の出力にしたがって行動させる．この時，4 つの出力のうち一番大きな値を 1 とし，その他の値を 0 としてそれに対応する行動をとる．例えば出力が(0.1, 0.1, 0.4, 0.8)であったとき，教師信号が(0, 0, 0, 1)である「前進」の行動をとる．

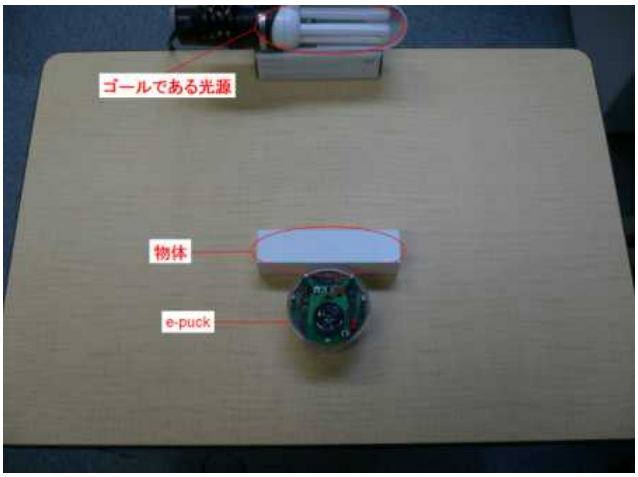

図 4 タスク 1 における e-puck および物体の初期位置

4.2 タスク 2

タスク 2 では図 5 のように e-puck と物体の延長線上にゴールがない環境で実験を行った. 図 5 に示したように, 回り込むようにして物体を押してゴールまで運ぶように操縦した. 5 試行分の学習データを用意し, 単体 NN と ME を用いて学習を行った. 5 試行中に得られた入出力回数は 117 回である. 単体 NN は入力層のノード数 12, 出力層のノード数 4 の 3 層 NN を用意し, 学習率は 0.1 とした. 中間層のノード数は 20 と 60 の 2 通り, 学習回数は 2000 回と 10000 回の 2 通りの合計 4 通りで学習を行った.

ME のエキスパートネットは 3 つ用意した. 各エキスパートネットには入力層のノード数 12, 中間層のノード数 20, 出力層のノード数 4 の 3 層 NN を使用し, 学習率は 0.1 とした. ゲーティングネットには入力層のノード数 12, 中間層のノード数 20, 出力層のノード数 3 の 3 層 NN を使用し, 学習率は 0.1 とした. 学習回数は 2000 回と 10000 回の 2 通りで学習を行った.

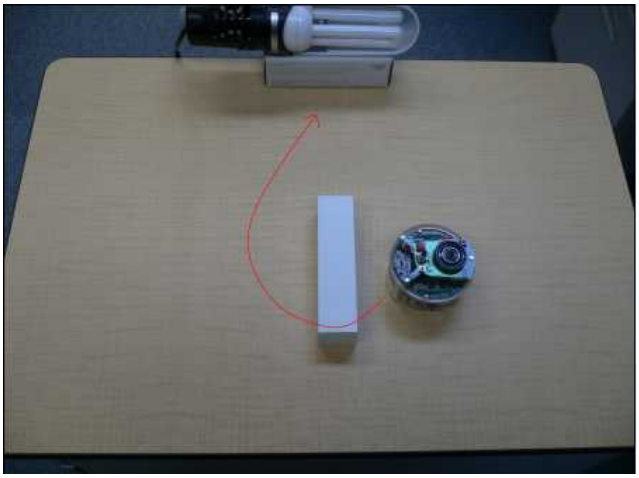

図 5 タスク 2 における e-puck および物体の初期位置

4.3 タスク 3

タスク 3 では図 6 のように e-puck の初期位置を 2 つ設定し, 複数の経路を学習させる. タスク 2 と同様に, 回り込むようにして物体を押してゴールまで運ぶように操縦した. タスク 2 で学習に用いた初期位置 1 からスタートした 5 試行分の学習データに加えて, 初期位置 2 から回り込むように物体を押してゴールまで運ぶように操縦した 5 試行分の学習データ, 合計 10 試行分の学習データを用意した. 10 試行中に得られた入出力回数は 233 回である. この学習データをもとに, ME および提案手法を用いてそれぞれ学習を行った.

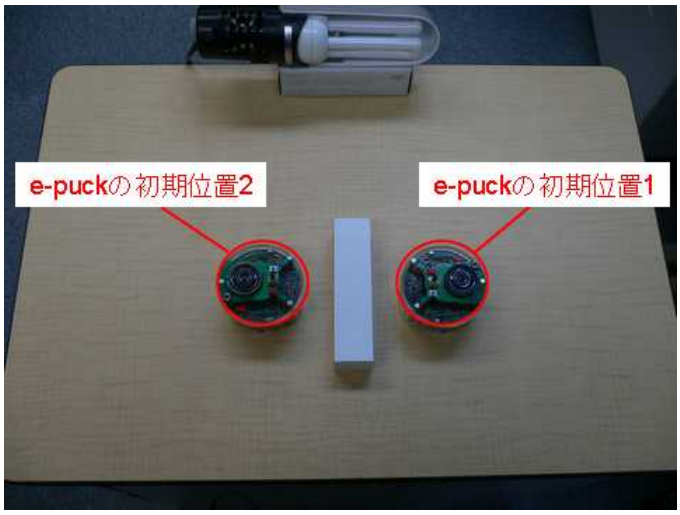

図 6 タスク 3 における e-puck および物体の初期位置

ME のエキスパートネットはタスク 2 と同様に 3 つ用意した. 各エキスパートネットには入力層のノード数 12, 中間層のノード数 20, 出力層のノード数 4 の 3 層 NN を使用した. ゲーティングネットには入力層のノード数 12, 中間層のノード数 20, 出力層のノード数 3 の 3 層 NN を使用し, 学習率は 0.1 とした. エキスパートネットの学習率は 0.1 と 0.3 の 2 通り, 学習回数は 2000 回と 10000 回の 2 通りの合計 4 通りで学習を行った. 提案手法の実装について以下に述べる.

[行動候補生成器について] 行動候補生成器には同タスクで学習を行った ME を用いた. また, 行動候補を決定する際の閾値を 0.4 とした. たとえば ME の出力が(0.1, 0.1, 0.6, 0.8)であったとき, 行動候補として教師信号が(0, 0, 1, 0)である「右折」と教師信号が(0, 0, 0, 1)である「前進」が行動候補として出力される.

[予測器について] 予測器にはエキスパートネットが入力層のノード数 16, 中間層のノード数 30, 出力層のノード数 12 である ME を用いた. エキスパートネットの学習率は 0.3, ゲーティングネットの学習率は 0.1 とし, 10000 回学習を行った. 予測器はある状態である行動をとると次にどのような状態になるかを学習する.

[状態評価器について] 物体と光源が目前にある状態をゴールとする.学習データからゴール状態における入力を推定し, その値を状態評価器でのゴール状態かどうかの判断に使用した.具体的には, 予測された図3の IR0, IR1, IR2, IR5, IR6, IR7 の赤外線センサと近接センサの正規化した値が以下の条件を満たせばこの状態をゴールであると判断する.

(1) **赤外線センサ**
- IR0> 1.2 かつ IR7 > 1.2
- IR2 < -0.8 かつ IR3 < -0.8

(2) **近接センサ**
- IR0 < -1.2 かつ IR7 < -1.2
- IR2 > 0.8 かつ IR3 > 0.8

これは e-puck の前進方向に物体があり、かつ光源に十分近づいた状態を表す.

[予測回数の上限について] 予測回数の上限は 25 ステップ先までとした.周囲の環境情報が入力され, 一回行動するまでを 1 ステップとする.25 ステップ先まで予測した結果, ゴールの状態を発見できなかった場合, 単体 NN や ME で行動を決定する際と同様に一番大きな出力に対応する行動をとる.

[予測回数の上限について] 予測回数の上限は 25 ステップ先までとした. 周囲の環境情報が入力され, 一回行動するまでを 1 ステップとする. 25 ステップ先まで予測した結果,ゴールの状態を発見できなかった場合,単体 NN や ME で行動を決定する際と同様に一番大きな出力に対応する行動をとる.

行動候補生成器には前述したエキスパートネットの学習率と学習回数を変えて学習を行った 4 種類の ME を用いることで,4 通りの実験を行った.

5. 実験結果および考察

それぞれのタスクについての結果および考察を述べる.

5.1 タスク 1 における実験結果および考察

初期位置からスタートさせて, 物体を押してゴールまで運ぶか物体が机から落ちてしまうまでを 1 回として, 10 回実験を行った. その結果, 10 回とも e-puck は物体をゴールまで押して運ぶことができた. 物体がずれるのに合わせて e-puck も位置を調整しながらゴールまで物体を運ぶ様子が見られた. 比較的単純な入出力系列であったため, 単体 NN でうまく学習できたと考えられる.

5.2 タスク 2 における実験結果および考察

タスク 2 における 10 試行中のタスク達成回数を表 1 に示す. 単体 NN では回り込みながら物体を押してゴールのほうに向けるという行動をとることができず, 直進し続ける様子が見られた. 原因として, タスクを達成するためには他の行動に比べて直進を多く行う必要があることが考えられる. 学習データでは 117 回の入出力が行われているが, そのうち直進が 80 回, 右折が 28 回, 左折が 9 回となっている. そのため, 特に左折がうまく学習できず, 回り込むという行動を学習することができなかったと考えられる.

ME を用いて学習することによって, 回り込みながら物体を押してゴールのある方向に向けたのち, 位置調整をしながらゴールまで物体を運ぶ様子が見られた. NN と比較して ME を用いると設定すべきパラメータが増えるが, 複雑な入出力系列が行われる環境では有用であることがわかる.

表 1 タスク 2 における実験結果

使用したアルゴリズム	学習回数	10 試行中のタスク達成回数
単体 NN (中間層のノード数 20)	2000	1
	10000	0
単体 NN (中間層のノード数 60)	2000	0
	10000	0
ME	2000	7
	10000	8

5.3 タスク 3 における実験結果および考察

タスク 3 における 10 試行中のタスク達成回数を表 2 に示す. エキスパートネットの学習率 0. 1 で 2000 回学習を行った ME では, 回り込みながら物体を押してゴールのほうに向けるという行動をとらずに直進し続ける様子が見られた. これはタスク 2 の単体 NN の結果と同じで, 学習データ中に直進の占める割合が他の行動と比べて多いことが原因として考えられる. しかし, 学習回数を増やすことでタスクを達成できた回数が増加している. 学習回数が増加することで, エキスパートネットの分担がうまく行われて学習が進んだと考えられる.

行動候補生成器にエキスパートネットの学習率 0. 1 で 2000 回学習を行った ME を用いた結果に着目する. ME だけで行動決定をした場合, 一度もタスクを達成することができなかったが, 予測情報を元に行動決定をすることで 8 回タスクを達成することができた. その 8 回の試行を観察すると, 初期位置からスタートしてゴールに物体を押して運ぶまで平均で 25 ステップ程度かかっていた. その際, 行動候補生成器の一番大きな出力に対応した行動とは異なる行動を予測に基づいて選択した回数は平均 7 ステップ程度であった. その全てが物体を回り込むように押してゴールのある方向に向ける過程で見られた. 位置調整をしながら物体をゴールまで押す過程では常に行動候補が 1 つしか生成されず予測は行われなかった. つまり, ME を用いた 2000 回の学習では, 位置調整をしながら物体を押すことは学習できているが回りこむように物体を押すことは学習しきれていないことになる. これは, 学習が十分にできておらず, 現在取るべき行動が絞りきれないときに, 予測情報に基づいて行動を決定することによって適切な行動が生成できることを示している.

表 2 タスク 3 における実験結果

使用したアルゴリズム		10試行中の達成回数	
		ME	提案手法
学習率0.1	学習回数2000	0	8
	学習回数10000	7	8
学習率0.3	学習回数2000	5	8
	学習回数10000	9	8

6. おわりに

本研究では, 繰り返し予測を行うことによって遠い未来の状態を予測し, その予測情報に基づいて行動を生成する手法を提案した. 実機実験に適用した結果, 現在取るべき行動が絞りきれない時に繰り返し予測を行うことによってタスクの達成率が上昇した. 予測情報を用いる利点として, ある状態で望ましい行動が複数ある場合, どの行動が本当に望ましいか先読みして決めることができる, という点がある. この利点を示すためには, タスク達成に複数の手段があるタスクで今後実験を行う必要がある. また, 今回行った実験では行動候補を生成する際の閾値は手動で設定した. 実環境下で活動する自律ロボットを実現するためには, この閾値も学習によって獲得することが望ましい. 学習データからの学習によって, 閾値が自動的に学習されるような手法を考える必要がある.

参考文献

[Tani 1999] J. Tani, S. Nolfi: ``Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems", Neural Networks, Vol. 12, pp. 1131-1141, 1999.

[Ohigashi 2003] Yu Ohigashi, Takashi Omori, Koji Morikawa, Natsuki Oka: ``Acceleration of Game Learning with Prediction-based Reinforcement Leaning -Toward the emergence of planning behavior-", ICANN/ICONIP 2003, LNCS 2714, pp. 786-793, 2003.

[Jacobs 1991] R. Jacobs, M. Jordan, S. Nowlan, and G.Hinton: ``Adaptive mixture of local experts", Neural Computation, Vol. 3, pp. 79-87, 1991.