複合的タスクにおける既存知識の選択規則の学習

識名 翔 † 服部 元信 ††
† 山梨大学大学院医学工学総合教育部　†† 山梨大学大学院医学工学総合研究部

1 はじめに

近年，科学技術の進歩により様々な分野においてロボットが活躍しており，その用途も掃除や警備，災害救助など多種多様にわたる．しかし，現存するロボットの多くは特定用途向けであり，その行動制御は設計者に依存する．そのため今後は複合的な環境や設計者が想定困難な環境など，より複雑な環境に対しても適応的かつ自律的に行動を制御するような高次な機能を持つロボットが期待される．

そのような人間や他生物と同様に自律的な行動制御が行えるエージェントを設計する場合，目標である人間や他生物の行動や学習の仕組みを模倣することは自然なアプローチである．

われわれ生物は新たな環境に直面した際，既に持っている行動を知識とし利用して対応する．例えばサッカーの「ドリブル」という行動を行う場合，よりプリミティブな「ボールを蹴る」,「ボールを追いかける」という2つの行動を状況によって使い分け「ドリブル」を行う．このように既存知識の組み合わせによる行動により，より複雑な行動を行うことができる．

そこで本研究では，複合的な行動が求められる環境においても複数の基礎的知識を状況に対して適応的に選択，行動するための選択規則獲得モデルを提案する．本研究で扱う知識とは，【入力】環境情報 → 【出力】行動，のような処理を行う手続き的知識のことを指し，これを階層型ニューラルネットワークにより構成する．

それにより，ある環境で獲得した選択規則も一つの知識と見ることができ，また別の環境で利用することが可能で学習時間の短縮等が期待できる．

学習モデルの可能性を測るため，最初は単純な環境でプリミティブな行動での選択規則を学習，次にそこで獲得した選択規則を既存知識として加え，より複合的な環境で学習といったように，段階的に複雑な知識を獲得していき，その有効性を検証する．検証には実環境とのインタラクションが行え，環境に働きかける行動ができる実ロボットが適していると考えられる．

そのため，移動ロボットと，そのシミュレートを行うシミュレータを用いて，計算機シミュレーションと実環境実験の両面でその有効性を検証する．

実環境で学習を行うには膨大な時間を要するため，本研究では計算機シミュレータを用いて，ノイズも含めた実環境を再現したシミュレーション環境において学習を行わせ，その結果を実環境に適用する方法をとる．

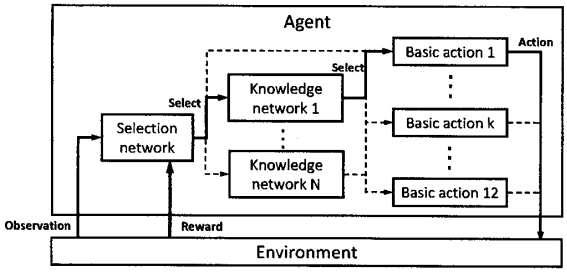

図 1: 既存知識を用いた選択規則獲得モデル

2. 既存知識を用いた選択規則獲得モデル

本研究では，人間の行動獲得の過程を模した学習方法として，複数の既存知識の協調行動により，それらを複合した知識選択ネットワークの獲得を行う学習モデルを提案する．

図1はこの学習モデルの一般的な形となる．まず環境の状態が知識選択ネットワークに入力され，知識1～Nのニューラルネットワークまたは後述する12種の基本行動知識のどれを用いるかが選択される．そして，基本行動知識が選択された場合は対応する行動を，知識1～Nが選択された場合は選択されたニューラルネットワークに環境の状態が入力され，その出力により再度基本行動知識の選択が行われ，行動が行われる．

知識選択ネットワークでは既存知識の数と同じ数の出力層ニューロンを用意し，出力値を各知識の行動価値として扱う．そしてネットワークの計算で得られた行動価値の各々に微小乱数を付加し，確率的に知識選択を行う．そうすることで行動価値の高い既存知識が優先されると共に，低い知識もある程度確率的に選択される．乱数要素は学習回数と共に小さくしていく．学習時は，選択した既存知識に該当する出力のみ教師信号を与えて誤差逆伝播（Back Propagation：BP）法により学習させる．

これにより知識選択ネットワークは環境の状態に適

Learning for selection of existing knowledge in complex tasks
† Sho SHIKINA
†† Motonobu HATTORI
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (†)
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (††)

当な既存知識の選択規則を学習し，単体の知識のみでは困難であった環境にも適応することが可能となる．

3. 計算機シミュレーション

提案法の有効性を検証するため，まずロボットシミュレータ Webots を用いて移動ロボット Khepera（図 2 a 参照）の計算機モデルに提案学習モデルを実装し，計算機シミュレーションを行った．

実験を始めるにあたり，まずロボットのプリミティブな知識として図 2 b に示す 12 種の基本行動知識を規定した．この 12 種の知識選択により移動ロボット Khepera の全動作が行える．また，ロボットの入力には Khepera が装備する 1 次元 64pixel のグレイスケールカメラおよび周囲 8ヵ所にある赤外線・光を感知する各センサからの出力値を使用し，Direct-Vision-Based 強化学習 [1] により学習を行った．

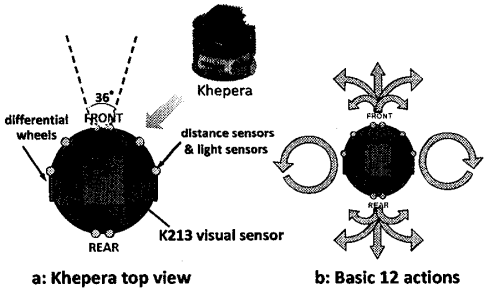

図 2: a:Khepera 上面図，b:12 種の基本行動

実験の第 1 段階として，単純な環境での選択規則の学習を行った．扱うタスクとしては「目標物に向かって進む」行動が求められる『目標物到達』と「光に向かって進む」行動が求められる『光到達』，そして「障害物を避けながら進む」行動が求められる『障害物回避』をそれぞれ 12 種の基本行動知識を用いて選択規則を獲得させた．

次に第 2 段階として，第 1 段階で獲得した 3 種の選択規則を既存知識として新たに加え，15 種の既存知識を用いてより複合的な環境での選択規則の学習を行った．扱うタスクとしては，「障害物をかわしながら光へと向かっていく」行動が求められる『障害物回避光到達』と「障害物をかわしながら目標物へと向かっていく」行動が求められる『障害物回避目標物到達』の選択規則の獲得を行った．図 3 に各タスクにて獲得される行動のイメージを示す．

学習環境は周囲を壁に囲まれた箱庭である．目標物・光源に到達するか，障害物に衝突するか，規定時間に達するかを 1 試行として学習を行い，毎試行ごとに，目標物・光源は Khepera の初期位置から一定の範囲でランダムに，障害物は 10 個の立方体をランダムな

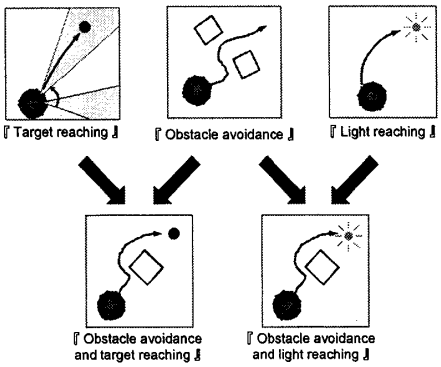

図 3: 各タスクにおける行動のイメージ

位置に配置した．

この環境で第 2 段階のタスクを行ったところ，獲得までの試行回数は『障害物回避目標物到達』が 3210 回『障害物回避光到達』が 2900 回（共に 30 シミュレーション平均）であったが，Direct-Vision-Based 強化学習により 12 種の基本行動のみで同タスクを学習させたところ，10 万試行を上限として学習ができず，学習困難であると判断した．これにより，このような複合的な知識を一から学習するのは困難であり，提案法のような手法によって獲得することが望ましいと言える．

4. 実環境実験

計算機シミュレーションでの結果を元に実環境での『障害物回避目標物到達』および『障害物回避光到達』タスクの検証を行った．環境設定は計算機シミュレーション時と同様である．

実験の結果，ノイズの影響のためか計算機シミュレーションと比較すると行動選択にブレが生じるものの，両タスクともゴールまで達成することができ，提案モデルの有効性が確認できた．

5. まとめ

複数の既存知識を利用し状況に応じて使い分ける選択規則を獲得することで，より複合的な環境にも適応できる学習モデルを提案した．実験により簡単なタスクで獲得した選択規則は新たな既存知識として利用でき，さらに複合的なタスクでの選択規則も獲得することができた．今後はより複雑な環境での選択規則の獲得を行い，既存の知識を使い分けることでどの程度まで複雑な選択規則を獲得できるかの検証を行っていく．

参考文献

[1] 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.