1F1-01

非マルコフ問題解決のための適応型 XCSM（A-XCSM）の提案

Adaptive XCSM for perceptual aliasing problems

劉 淑梅[*1]
Liu Shumei

長尾 智晴[*1]
Nagao Tomoharu

[*1] 横浜国立大学
Yokohama National University

Recently, many works about perceptual aliasing problem has attracted increasing attention in autonomous agent control. In XCSM, by adding constant length of memory to general XCS, the agent gets suitable action, combining the present information with the past information. But its performance is decreased because of the fixed-length memory. In this report, an adaptive XCSM method has been proposed by involving a changeable length of memory. Using Q-learning method, suitable memory length for each rule can be obtained effectively.

1. はじめに

エージェントの行動制御におけるエイリアス問題に対しては、エージェントの履歴を考慮し、行動を選択する必要がある。XCS を用いたエージェントの行動制御において、これを実現するため、Pier Luca Lanzi は XCSM（XCS with Memory）を提案している。XCSM ではエージェントが利用できる内部状態数を固定して問題に適用する。しかし、この内部状態数はあらかじめ設定しなければならず、XCSM の性能は固定した状態数に左右される。

そこで、本報告では、進化過程で扱う状態数を増減させる適応型 XCSM（A-XCSM）を提案する。本手法は問題解決に必要な状態数を進化により自動的に獲得するため、未知の環境に対しても状態数に対する試行錯誤を必要としない。A-XCSM の性能を検証するために迷路探索問題に適用し、従来の XCSM との比較を行ったので報告する。

2. 適応型 XCSM（A-XCSM）

2.1 XCS

XCS とは、「if<条件> then <行動>」形式のルール集合を用いて、外部環境からの入力に対してある行動を出力するシステムである。その条件部は {0,1,#}、行動部は {0,1} の集合で表される。"#"は 0 でも 1 でもよい"do not care"を表している。ルールの定義は次のように行われ、状態部と条件部以外はルールを評価するための適応度がある。

ルール： {条件部、行動部、適応度、…}

XCS の枠組を図 1 に示す。

詳しいアルゴリズムの各ステップを述べる。

- 初期個体集団の生成
- マッチング集合の生成

環境から取得した状態に基づいて、個体集団からマッチングしたルールを選択し、マッチング集合（mSet）を生成する。mSet には、各動作に対して最低一つのルールが包含される。包含されない動作があれば、covering を利用して、新しいルールを作成する。mSet のルールに対して、平均的な適応度が最大の動作を選択し、環境に作用して、報酬を得る。

- 動作集合の生成と更新

mSet に選択した動作をもっているルールから動作集合（aSet）を生成する。この動作集合に対して、similar-Reinforcement Learning を用いて、各ルールの適応度などのパラメータを更新する。次に GA を利用して、新しい個体を生成する。

- 異種個体の削除

個体集団が際限なく膨張することを避けるために、ランダムに選んだルールを一つ削除する。

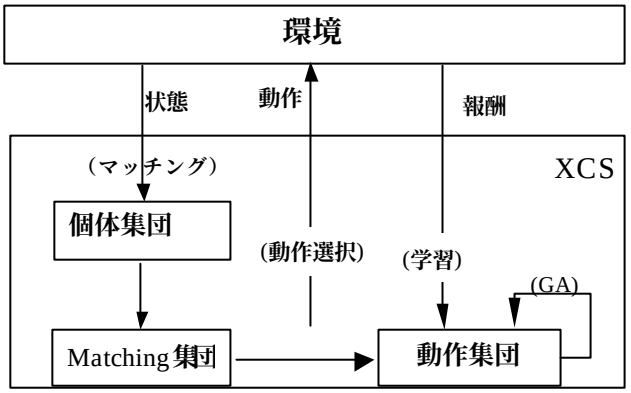

図 1 XCS の枠組

2.2 XCSM

XCS は、マルコフ問題に対して有効性が高いことが既に証明されている。しかし、エイリアス問題に対しては、あまり有効ではない。このような問題に対して、XCS に内部メモリーを加えた XCSM[(1)]が提案された。

XCS ではエージェントは環境から得た状態以外に、内部レジスタをもつ。ルールには、同じの長さのシンボルをもつ内部状態と内部動作が加えられた。内部状態はレジスタとマッチングされ、内部動作はこのレジスタを修正する。内部状態のシンボルは普通の状態と同じ意味をもっている。内部動作には、0 と 1 の場合に内部レジスタにも 0 と 1 に設定し、＃の場合は内部レジスタを変更しない。ルールは次のように定義される。

ルール： {条件部、行動部、内部条件、内部動作、適応度、…}

XCSM の処理の流れは XCS と比べて、二つの違いがある。一つ目はマッチング集団を生成するとき、環境から得た状態と内部レジスタを一緒にマッチングすることであり、二つ目は最良の動作を選択するとき、動作と内部動作が並行的に処理されることである。

[*1] 横浜国立大学 大学院環境情報学府
Graduate School of Environment and Information Sciences, Yokohama National University

しかし、内部動作と内部状態はmビットを設定する場合は、可能的内部動作数は最大 3^m である。mが大きくなると、動作の探索空間は膨大になる。この欠点を解決するために、適応型 XCSM（AXCSM; Adaptive XCSM）を提案する。

2.3 AXCSM

AXCSM のルールでは、内部条件と内部動作の長さは不確定であり、人間があらかじめ設定する最大値まで、ランダムな何段階かの長さがある。あるルールにどの長さが最適かを確定する方法は AXCSM の最も重要な点である。本報告で Q-Learning 方法を利用して、内部動作の最適な長さを学習する。

AXCSM のルールは次のように定義される。

ルール： {条件部、動作部、内部条件、内部動作、内 部動作の長さ、適応度、…、Q-Value}

Q-Value は内部動作の長さを評価するためのパラメータである。

最適な長さを決定する枠組を図２に示す。まずマッチング集合の各動作ごとに、内部動作の長さによりグループに分ける。各グループごとに Q-Value 値を平均し、平均値が最大の長さをこの動作の最適な内部動作の長さと決定する。次に、m Set から内部動作の長さと一致しないルールを削除し、m Set を更新して、新しいマッチング集合を決定する。

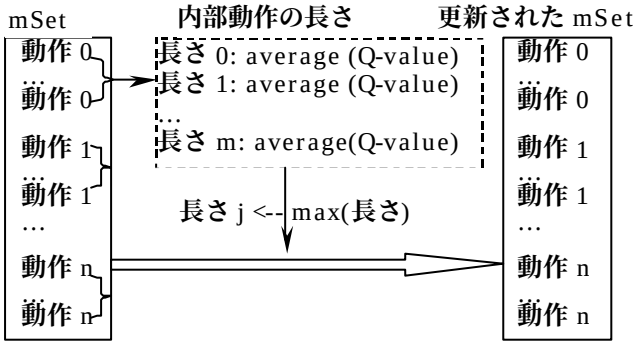

図２ AXCSM で内部動作の長さを決定する枠組

最適な長さを決定した後は、動作集合が学習するとき、次の公式に基づくQ-Learning 方法を利用して、Q-value が学習される。

$$Q(s_t,a_t) += \alpha\left[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right]$$

(公式 1)

この公式において、s_t は時刻 t における選択したルールの動作、a_t は内部動作の長さを示す。r_{t+1} は環境から得た報酬である。a と? は計算パラメータである。

3. 迷路探索問題の設定

本報告における迷路探索問題を次のように設定する。

- エージェントの視野は3×3である
- エージェントの行動は周囲 8 方向（N, EN,E,ES,S,WS,W,WN）である
- スタート位置は毎回ランダムに設定され、ゴールは迷路に"G"と表される
- エージェントの学習目標は、任意の位置からスタートして、ゴールまでの最短経路を探索する事である

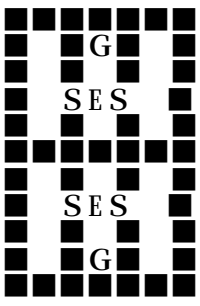

図３ 迷路の例

迷路は図 3のような 7×11 のフィールドを用意する。この迷路に S と表示されている状態は、位置が違い、適切な動作も各々違う。しかし、エージェントの視野内の情報が同じであるので、適切な動作を選択することは困難である。このような状態をエイリアス状態という。E も同様なエイリアス状態である。

4. 結果

図 4および図５に、ゴールまでの平均ステップ数を 10 試行の結果の平均として示す。

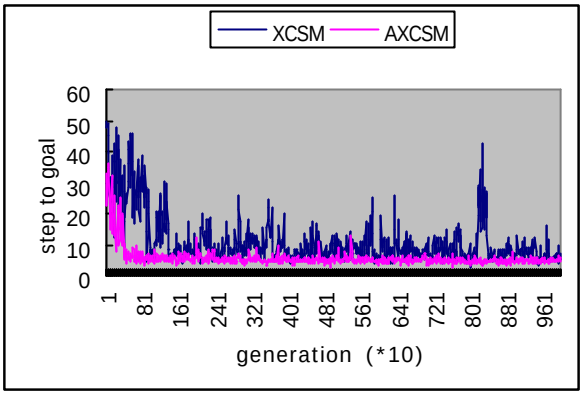

図４ ゴールまでのステップ数(内部動作 3bit)

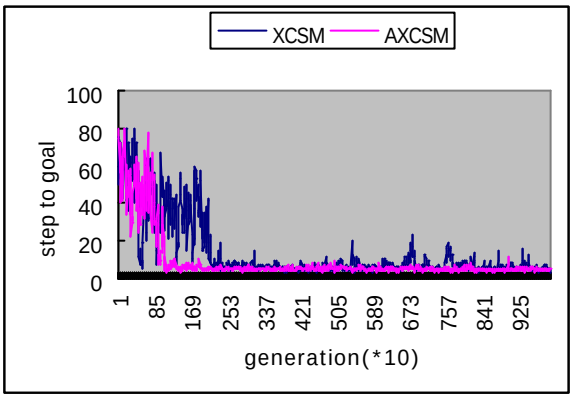

図５ ゴールまでのステップ数（内部動作 4bit）

図４，５から、本手法は実験した迷路に対して、より早く収束し、安定になったことが分かる。このことは、本手法が、同一の入力に対して異なる出力を必要とする問題に対して有効であり、また、収束も早いことを示している。

5. 終わりに

本報告では、適応型 XCSM を提案し、エイリアス問題に対して適用して有効性を示した。今後は、より多くの状態数を必要とする問題に対して本手法を適用するとともに、今回のような静止的な環境ではなく、動的な環境での実験を行う予定である。

参考文献

[Lanzi 2000] Pier Luca Lanzi, Stewart W.Wilson: Toward Optimal Classifier System Performance in Non-Markov Environments, Evolutionary Computation 8(4); pp393-418, 2000

[福寄 2002] 福寄雅洋,原 章, 長尾智晴 :不完全知覚問題解決のための時系列依存分類システム（TCS）の提案 電気学会論文誌 C, Vol.122-C,No.7,pp1218-1225,2002