2L5-4

クラスタリングを用いたクラシファイアシステムの改良

Improvement of the Classifier System Using Clustering

林田智弘 Tomohiro Hayashida　西崎一郎 Ichiro Nishizaki　関崎真也 Shinya Sekizaki

広島大学
Hiroshima University

In many POMDPs (Partially Observable Markov Decision Process), there exist the problem of aliased states such that the appropriate action cannot be chosen based only on the current environmental information that an agent observes. Several classifier systems have been developed to avoid such problems through including internal memory to register past information of observed environments and chosen actions. However, the internal memory sometimes requires an enormous amount of memory area. This paper suggests an effective method for utilization of the internal memory through a application of clustering method, and improves a classifier system, ACSM (Anticipatory Classifier System with Memory) which had been constructed by Hayashida et al. (2014).

1. はじめに

「もし〜ならば〜する」というような if-then の関係にある条件部と行動部によって構成されるルールをクラシファイアと呼び，クラシファイアの集合をクラシファイアシステムと呼ぶ．クラシファイアシステムは，if-then ルールに従い環境からの入力に対して行動を出力し，さらにそのルールが適切であったかどうかを学習することでよりよい方策を獲得することができる．

これまで，部分可観測マルコフ決定過程 (POMDP: Partially Observable Markov Decision Process) のベンチマークである迷路問題を用いて，様々なクラシファイアシステムや，その他の強化学習に基づくシステムの性能評価が行われてきた．

XCS(eXtended Classifier System)[4] は，予測獲得報酬の予測精度に基づいてルールを更新するクラシファイアシステムであり，ACS2 (Anticipatory Classifier System 2)[5] は，エージェントの行動による環境の変化の予測精度に基づいてルールを更新するクラシファイアシステムである．しかし POMDPs には，システムが認識する環境に対して最適な行動を一意に決定できないエイリアス状態が存在する．これに対して，過去の状態や行動を一時的に保持する XCSAT(XCS with an internal Action Table)[2] や, ACSM(ACS2 with Memory)[1] などが提案されている．エイリアス状態を含む迷路問題では，ACSM が従来研究で提案されたクラシファイアシステムの中で最も有効であることが示されている [1]．特に ACSM は，迷路問題より一般的な最短経路発見問題においても高いパフォーマンスを示し，一方で，内部メモリを効率的に取ることができず非常に実行時間がかかることが示された [3]．

本研究では，ACSM の内部メモリをクラスタを用いて改良し，POMDP 問題である迷路問題や最短経路発見問題において，従来のクラシファイアシステムと比較検証を行う．

2. 従来研究

2.1 HayasahidaEtal14

ACSM[1] は，POMDPs におけるエイリアス状態に対応するために，ACS2 に内部メモリを追加したものである．ACS2 は，まずエージェントが環境を認識し，条件部が環境と一致するクラシファイアの集合であるマッチセットを構成する．マッチセットの中から，予測精度 p と予測報酬 r の積である適合度 $F = p \times r$ に基づいて行動を選択し，マッチセットに含まれるルール群のうち，選択された行動と同じ行動部を持つルール群によりアクションセットを構成する．エージェントは選択されたルールに従って行動し，アクションセットのクラシファイアは環境から報酬を得る．さらに，ALP(Anticipatory Learning Process) を用いて環境変化をもとに予測精度を更新し，報酬に基づき Q 学習を用いて予測報酬を変更し，遺伝的アルゴリズム (GA) を適用して学習を行う．

ACSM では，ある環境に対応したマッチセットにおいて，行動部が異なり，かつ適合度が同程度に高いクラシファイアが含まれていれば，エージェントは最適な行動を選択できないため，現在の環境をエイリアス状態と判定する．このとき，エイリアス状態に入る前の状態と，現在までの相対座標を内部状態に記録しておくことで，エイリアス状態を判別し，これにより最適な行動が選択可能となると考えられる．

ACSM の概要を図 1 に示す．

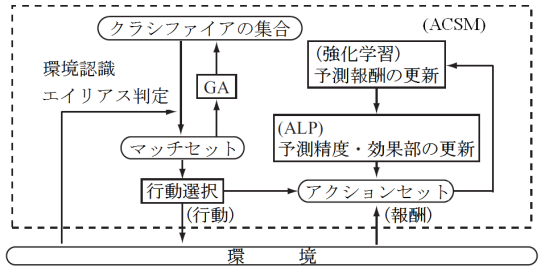

図 1: ACSM の枠組み

2.2 POMDP

MDP(Markov Decision Process) とは，次状態への遷移が現在の状態のみに依存して決定する確率的逐次決定過程である．POMDP とは，MDP のうち，エージェントが部分的な環境情報しか認識できない状況である．多くの POMDP はエ

連絡先: 林田智弘, 広島大学, 東広島市鏡山 1-4-1, 電話: 082-424-5267, E-mail: hayashida@hiroshima-u.ac.jp

イリアス状態を含み，このため if-then 形式の行動ルールでは最適方策の獲得が難しい．

本研究で取り扱う POMDP 問題は，迷路問題と短経路発見問題とし，これらの問題において ACSM の効率的なメモリ確保を行うシステムの開発を目指す．

2.2.1 迷路問題

迷路問題は，POMDP 問題における基本的なベンチマーク問題であり，エージェントはスタートからゴールへ短いステップ数での到達を目標とする．最も単純な問題例として図 2 の woods100 が挙げられる．図 2 の白色のセルは通路であり，黒色のセルは壁であり，G は目的地のゴールを示す．迷路問題において，エージェントは現在地点の周囲 8 セルの通路・壁・ゴールを知覚し，それを条件部とする．例えば，セル 1 においてエージェントは右のセルのみ通路を知覚し，それ以外のセルは壁を知覚する．セル 1 では右に進むことが最適となる．ここで，セル 2 とセル 5 に注目すると，共に左右を通路と知覚する一方，最適な行動がセル 2 では右，セル 5 では左となる．このように，同じ条件部を持ち，最適な行動が異なる状態をエイリアス状態と呼ぶ．ACSM では，エイリアス状態に入る前の知覚情報と現在地点までの相対座標を内部メモリとし学習を行う．

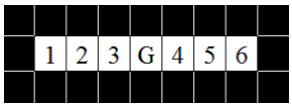

図 2: 迷路問題 (woods100) の例

2.2.2 最短経路発見問題

最短経路発見問題とは，迷路問題を一般的に拡張した問題である．最短経路発見問題の例を図 3 に示す．ここでは，エージェントが一度に知覚できる範囲は図 3 に示すように，周囲 8 方向の 3 セル先までとし，知覚できる範囲内でエージェントは行動を決定する．なお，壁の先はエージェントは知覚できないため，壁と判断する．最短経路発見問題では，迷路問題と同様にエイリアス判定を行う．しかし，最短経路発見問題では，最適な経路が複数存在する場合，システムがエイリアス状態と認識する状態が無駄に増加する．例えば，図 3 に示される経路発見問題では，エージェントはゴール G に移動したいとき，最短経路は，{ 右 (d_1)3 セル → 右上 (d_2)1 セル } 移動する経路と，{ 右上 (d_2)1 セル移動 → 右 (d_1)3 セル } 移動する経路の 2 つ存在する．このとき，エージェントの現在の地点において，d_1 に 3 セル移動するクラシファイアと，d_2 に 1 セル移動するクラシファイアの適合度に差がなく，本来エイリアス状態ではない環境であっても誤ってエイリアス状態と判定してしまう可能性がある．そのため，マッチセットの中で同程度に適合度が高いクラシファイアが存在し，行動が互いに異なると判定する条件を，d_1 と d_3，または d_1 と d_7 のように 90° 以上離れている場合とする．

ACSM[1] は，最短経路発見問題において，問題の規模が大きくなるほど不必要な内部メモリを持つクラシファイアを学習の段階で多く生成するため，実行時間が長くなる問題がある [3]．本研究では，過去の環境情報や行動履歴などに関する情報を圧縮することでこの問題を解決できると考え，このような情報をクラスタリングにより情報量を圧縮し，これを ACSM の内部メモリに記録することで，システムの改善を目指す．

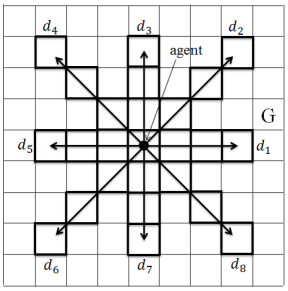

図 3: 最短経路発見問題の例と知覚範囲

3. ACSM のクラスタを用いた改良

本研究では，エージェントの各状態遷移をラプラシアン行列に変換し，スペクトラルクラスタリングにより各状態のクラスタ化を行い，そのクラスタ情報を ACSM の内部メモリに用いることで情報の圧縮を行う．

3.1 スペクトラルクラスタリング

従来研究では，MDP 問題に対しスペクトラルクラスタリングを用いて分析を行った報告がされている [6]．スペクトラルクラスタリングとは，ラプラシアン行列の固有値固有ベクトルを用いることでクラスタ化を行う手法であり，最小の固有値を持つ固有ベクトルによる分割が最適となる利点を持っている [6]．

ラプラシアン行列 $L = [l_{ij}]_{n\times n}$ とは，状態数 n としたとき，各状態間の重みからなる重み行列 $W = [w_{ij}]_{n\times n} \geq 0 (i, j = 1, \cdots, n)$(非負の対称行列) と各状態の隣接する重みの合計を対角成分に持つ定量行列 $D = [d_{ij}]_{n\times n} (d_{ii} = \sum_{j=1}^{n} w_{ij})$(対角行列) からなり，次の式 (1) を満たす．

$$L = D - W \tag{1}$$

ラプラシアン行列 L の固有ベクトルを x としたとき，固有値は以下の式 (2) を満たす．

$$x^t L x = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (x_i - x_j)^2 \tag{2}$$

式 (2) の右辺から，ラプラシアン行列の固有値は非負であることがわかる．また，重みの大きい状態間において固有ベクトルの要素の値が近い場合では $w_{ij}(x_i - x_j)^2$ の値は小さくなり，重みの小さい状態間では固有ベクトルの要素の差は大きくとも $w_{ij}(x_i - x_j)^2$ の値は小さくなる．つまり，$w_{ij}(x_i - x_j)^2$ の合計を小さくすることは，最適な分割を示すこととなり，それは，最小の固有値を求めることと同義となる．そのため，最小の固有値を求め，その固有ベクトルを用いることでクラスタ化が可能となる．

ここで，固有値 0 の場合について考える．$d_{ii} = \sum_{j=1}^{n} w_{ij}$ の定義から固有値 0 は必ず存在し，同じく定義から固有ベクトルの要素は 1 か 0 で構成されることがわかる．ある状態同士が，別の状態間を含めたの全ての遷移を通して連続している場合，その全ての要素は 1 となり，連続していない場合は 0 となる．このように完全に分割できる場合は固有値 0 を用いて分割が可能である．しかし，本研究で取り扱う問題の状態は連続しており，固有値 0 による分割は不可能であるため，0 以外の最小の固有値・固有ベクトルを用いてクラスタ化を行う．

さらに，ラプラシアン行列 L は対称行列であり，かつ固有ベクトル $\mathbf{1}$ を持つことから，固有ベクトル x は 1 と垂直であることがわかる．よって，固有ベクトル x の各要素は正と負を持ち，各要素の合計は 0 となる．本研究では，スペクトラルクラスタリングによる状態の正と負による 2 分化を反復的に用いることで行う [6]．このとき，異なる状態数を一番多く含むクラスタを反復的に処理することで，任意のクラスタ数まで分割を行う．

また，本研究では，内部メモリを用いない事前学習期間での各状態間の状態遷移回数を遷移行列として記録する．遷移行列は対象行列ではないため，遷移行列とその転置行列を足すこと [8] で状態間の重み行列として扱うものとする．

MDP 問題では，エイリアス状態が含まれず，各状態は全て異なった条件部を持つため，クラスタによる分析は容易である．しかし，POMDP 問題においてクラスタ化した場合，エイリアス状態を含むため，図 4 のクラスタ C_2 のように隣接していない地点においても同じクラスタであると判断されることがある．

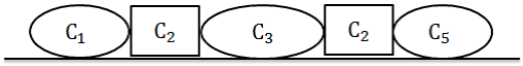

図 4: POMDP 問題におけるクラスタ化

ここで，迷路問題や最短経路発見問題において，クラスタ情報を効果的に用いるために，内部メモリには，ひとつ前のクラスタ番号と現在のクラスタに移った際の行動方向を用いることとする．図 4 を例にあげると，エージェントはクラスタ C_1 から右に移動しクラスタ C_2 に移動し，クラスタ C_2 内を移動後，クラスタ C_2 内でエイリアス状態の知覚情報を得たとする．このとき，条件部は，現在の知覚情報・ひとつ前のクラスタ番号 (C_1)・クラスタ遷移時の移動方向 (右) となる．ACSM とクラスタを用いたクラシファイアシステムのエイリアス状態における条件部の違いは，表 1 のように要約される．

表 1: エイリアス状態に対応するクラシファイアの条件部

ACSM[1]	現在の知覚情報 直近のエイリアス状態でない知覚情報 現在との相対座標
提案手法	現在の知覚情報 ひとつ前のクラスタ番号 クラスタ遷移時の移動方向

4. 数値実験

本研究では，迷路問題 (図 5)，最短経路発見問題 (図 6) を用いて提案手法であるクラスタを用いたクラシファイアシステムと，ACSM[1]，XCSAT[2] による実験を行う．なお，クラスタ数の上限は，迷路問題では 18，最短経路発見問題では 25 とした．

図 5 に示す迷路問題では，4 点のセル S のいずれかをランダムに選択してこれをスタート地点とし，エージェントは周囲 8 セルの壁・通路・ゴールを知覚し，ゴール G を目指して移動する．エージェントの 1 回の行動を 1 ステップとし，スタート

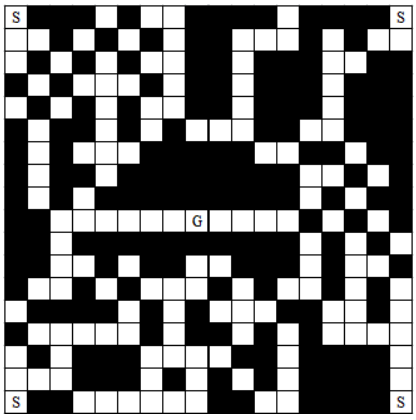

図 5: 迷路問題 (Large maze)

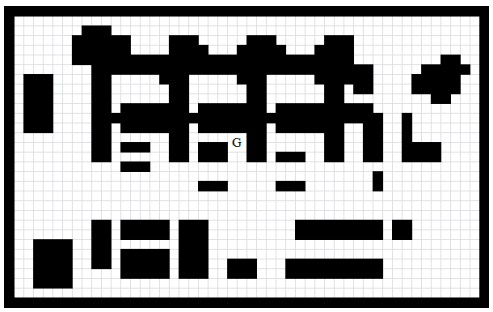

図 6: 最短経路発見問題

からゴールまでを 1 エピソードとする．行動選択として，確率 ϵ でランダムな行動を行い，確率 $1-\epsilon$ でマッチセットの中から適合度が最大となるクラシファイアの行動を選択する，ϵ-グリーディ手法を用いる．迷路問題では総エピソード数を 10000 とし，6000 エピソードまでは $\epsilon = 0.3$，それ以降は $\epsilon = 0$ とする．新たなクラシファイアが生成される場合のドント・ケアの発生割合 $p_\sharp$ について，3001–4000 エピソードでは $p_\sharp = 0.2$，それ以外は $p_\sharp = 0$ とする．また，2000 エピソードまでを提案手法および ACSM では ACS2 を用いて，XCSAT では XCS を用いた事前学習期間として，その後 9000 エピソードまでを各クラシファイアシステムによる本学習を行い，9001-10000 エピソードの期間は学習を行わず試行期間とし，この期間における 30 試行の平均ステップ数を実験結果とする．また，学習率 $\beta = 0.2$ とし，割引率はまっすぐ進んだ場合 $\gamma_1 = 0.7$, 斜めに進んだ場合 $\gamma_2 = 0.6$ と不要な斜め移動を防ぐためにバイアスをかけている．迷路問題 (Large maze) を用いた実験結果を表 2 に示す．

表 2 より，提案手法と ACSM には大きな差はなく，内部メモリに蓄積する情報をクラスタリングすることでパフォーマンスが落ちていることがわかる．これは，Large maze 程度の問題規模では，過去の情報を圧縮しないほうがむしろ効果的に内部メモリを活用できるといえる．

次に，最短経路発見問題を用いた実験では総エピソード数を迷路問題を用いた実験の 5 倍の 50000 としており，これに対応して ϵ-グリーディ手法における ϵ を変更する期間，事前学習などの期間，$p_\sharp$ を変更する期間などの値を迷路問題を用いた

表 2: 迷路問題 (Large maze) 結果

	提案手法	ACSM[1]	XCSAT[2]
平均ステップ数	14.2	**13.8**	21.1
最小値	**13.5**	**13.5**	16.1
最大値	15.5	**14.6**	24.1
実行時間	**27.7**	40.3	39.0

数値実験の 5 倍としている．図 6 に示す最短経路発見問題の白色のセルから 1 つをランダムに選択しこれをスタート地点とする．また，最短経路発見問題には多くの学習期間を要するため，学習率を $\beta = 0.2$ としている．最短経路発見問題を用いた実験結果を表 3 に示し，各項目について 3 種類の実験のうち最も高いパフォーマンスといえる結果が太字で示されている．

表 3: 最短経路発見問題結果

	提案手法	ACSM[1]	XCSAT[2]
平均移動距離	**30.8**	31.8	38.8
最小値	**28.1**	28.6	34.2
最大値	**34.6**	54.7	53.6
実行時間 (秒)	1538	5095	**1200**

表 3 から，提案手法は最短経路問題のような多くの学習が必要な問題に対してはパフォーマンスの向上が観測された．これは，迷路問題では，厳密な学習が必要であるため，クラスタを用いた場合は ACSM のピンポイントな学習に劣り，厳密な学習が難しい最短経路発見問題ではクラスタによる情報の圧縮により学習がうまくいったためであると考えられる．また，クラスタを用いたシステムは，内部メモリの情報の圧縮のために，実行時間に関しても性能向上が見られた．クラスタを用いたシステムの実験成功回数が多い理由として，厳密な学習を行わないために，パフォーマンスは悪くなっても最終的に柔軟に方策を得ることができたためであると考えられる．

5. まとめと今後

本研究では，ACSM の内部メモリにクラスタを用いたシステムを構築し，迷路問題及び最短経路発見問題に適用した．数値実験では，クラスタを用いたクラシファイアシステムを従来の ACSM と比較すると，迷路問題において良いパフォーマンスを得ることはできなかったが，より一般的な POMD P問題である最短経路発見問題においては優れていた．

今後は，迷路問題に特化したクラシファイアシステムである ACSST では対応できないような確率性を含んだ問題への適用が考えられる．

参考文献

[1] 林田, 西崎, 酒戸, “エイリアス状態を含む部分可観測マルコフ決定過程のための内部メモリを用いた ACS の開発,” 電子情報通信学会論文誌 A, J97-A, pp. 593–603, 2014.

[2] 西崎, 林田, 森分, “非マルコフ環境のための内部行動のテーブルをもつクラシファイアシステムの開発,” 電子情報通信学会 A, J94-A, pp. 982–990, 2012.

[3] 浅田, 林田, 西崎, “最短経路発見問題のための ACSM の改良,” *2013 IEEE SMC Hiroshima Chapter*, 若手研究会講演論文集, pp. 95–98, 2013.

[4] M.V. Butz and S.W. Wilson, “An algorithmic description of XCS,” *Soft Computing*, 6, pp. 144–153, 2002.

[5] M.V. Butz and W. Stolzmann, “An algorithmic description of ACS2,” *Advances in Learning Classifier Systems, 4th International Workshop*, pp. 211–229, 2002.

[6] N. Taghizadeh and H. Beigy, “A novel graphical approach to automatic abstraction in reinforcement learning,” *Robotics and Autonomous Systems*, 61, pp. 821–835, 2013.

[7] M.C.V. Nascimento, and A.C.P.L.F. Carvalho, “Spectral methods for graph clustering - A survey,” *European Journal of Operational Research*, 211, pp. 221–231, 2011.

[8] F.D. Malliaros and M. Vazirgiannis, “Clustering and community detection in directed networks: A survey,” *Physics Reports*, 533, pp. 95–142, 2013.