

北海道大学 ○ 三上 貞 芳, 嘉 数 侑 昇

## Abstract

相互に影響を及ぼしあう多数の独立したエージェントが、個々のエージェントでの学習により全体のパフォーマンスを向上するように進化する自己組織化モデルについて考察する。エージェントに確率的マトリックス学習オートマトン (Stochastic Learning Automata: SLA) を用いた場合に、この問題は多数の状態をもつ非定期的な反応を返す外部環境、すなわち他のオートマトンの状態に関する知識の獲得に帰着することを示し、その知識獲得の一方法として外部環境を有限個の状態ベクトルに代表させる自己組織化学習の導入を提案する。動的ジョブジョブスケジューリング問題へ適用し、シミュレーション実験を行ない、提案した構成の有用性を確認している。

## 1 はじめに

生産システムでのジョブ [UEDA91]・タスクスケジューリング [MIKA91]、自律移動ロボットのナビゲーション [SHIG91]などに代表されるように、確率的現象を入力とし、さらに多数のサブシステム (エージェント) が相互に影響を及ぼしあう複雑な系 (以下分散エージェントと呼ぶ) を対象として、系全体のパフォーマンスを向上させるように、エージェントの挙動をコントロールする問題は、今日の生産工学、ロボット工学を問わずさまざまな実問題を扱う上での重要な課題の一つとなっている。これに対しエージェント間の分散処理により、全体を安定状態すなわち目的の最適状態に到らせる自己組織化の考えが近年着目されている。しかし概念的議論はともかく、具体的に自己組織化の挙動を実現する各エージェントの構成については明示されていないのが現状である。

ここでは有限の行動をとる独立したエージェントを対象として、個々のエージェントでの学習により全体のパフォーマンスを向上するように進化する自己組織化モデルを提案する。エージェントに確率的マトリックス学習オートマトン (Stochastic Learning Automata: SLA) を用いた場合に、この問題は多数の状態をもつ非定期的な反応を返す外部環境、すなわち他のオートマトンの状態に関する知識の獲得に帰着することを示し、その知識獲得の一方法として外部環境を有限個の状態ベクトルに代表させる自己組織化学習の導入を提案する。

## 2 分散エージェント

分散エージェントとは、確率的挙動を示す外部系に作用して、目標を最適化する様に、インテリジェントなユニットである個々のエージェント  $a_i$  が独立して動作する系である。エージェント  $a_i$  は、他のエージェントの動作をも含めた外部系 (環境: environment) の状態  $S$  にしたがって、有限種類の動作 (行動:  $action_j$ ) から1つを確率的に選ぶものと一般化できる。また環境はエージェントの行動  $\{a_i\}$  を受け状態  $S$  を更新し、その結果に対する評価 (penalty)  $\beta$  をエージェントに返すものと一般化できる。

したがって系全体のパフォーマンス向上とは、環境からの  $\beta$  の期待値  $E(\beta)$  を大きくするように、各エージェントが自らの action の選択方法を変化させることと定義できる。

## 3 確率的学習オートマトンによるモデル化

もし (1) 環境の状態  $S$  の種類が有限で、かつ十分に小さく、(2) 環境が各状態  $S_k$  で定常的 (stationary) に反応する、すなわちエージェント  $a_i$  が action として  $a_j$  を取るときの penalty が

$$Prob(\beta|a_j^i) = c_{jk}^i = \text{const.}, \quad (1)$$

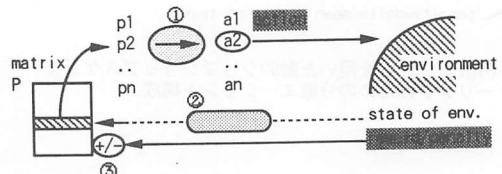
である場合、エージェントを確率的マトリックス学習オートマトン (SLA) とし、適切な再強化則 (reinforcement scheme) を導入することで  $E(\beta)$  の期待値が向上することが保障される [WITT77][NARE89]。

SLA は、 $k$  番目の状態  $s_k$  での  $a_j^i$  を選択確率  $p_{jk}^i$  にしたがって選ぶ確率システムである。  $P_k^i = p_{jk}^i$  は、  $a_j^i$  を選択した結果に対する penalty  $\beta$  にしたがって、再強化則

$$p_{jk}^i = p_{jk}^i - g_{jk}(P_k^i), \quad (2)$$

$$p_{j'k}^i = p_{j'k}^i + \sum_{j'' \neq j} g_{j''k}(P_k^i), \quad (3)$$

により再強化されるここで関数  $g_{jk}$  を  $g_{jk} = \alpha p_{jk}^i$  とした  $L_{R-I}$  スキームが  $\lim_{t \rightarrow \infty} E(\beta)$  を最適化する  $P_k^i$  へ収束させることが知られている。



- ① stochastic selection of an action.
- ② state categorization by unsupervised learning.
- ③ reinforcement of probability matrix P.

Fig.1 確率的マトリックス学習オートマトンと状態カテゴリー化学習の融合によるエージェントのモデル化。

しかし上記の前提に対して、実際に扱う問題では (1) 環境はエージェントの状態をも含めた複雑な挙動を示すゆえ、非定常 (non-stationary) と考えられ、(2) 環境の状態は多数の連続値変数を含む故に一般に少数の有限個とはならない。(1) に対しては、実験的に非定常環境に対しても再強化則により良好な  $E(\beta)$  が得られることがわかっている。とくに  $SL_{R-I}$  スキームが有効であると示されている [NARE89]。ところが (2) に対しては SLA はそのまま適用できず、したがって問題の状態空間を有限状態へ代表させる方法が必要となる (Fig.1)。

## 4 状態に関する知識の自己組織的獲得

$P_k^i$  の定義から SLA は状態  $S_k$  それぞれに対して独立した戦略をもつて行動する系とみなすことができる。この場合戦略とは行動選択確率ベクトルを意味する。したがって状態  $k$  とは各戦略が有効である環境

の状態を被うものであり、環境に関する知識の形成にほかならない。また SLA の性質から、同一の状態  $k$  での環境の反応は定常的であることが望ましいといえる。このような知識の獲得方法にはさまざまな考えられる。一例として後述する競合学習則による自己組織的カテゴリー形成では、状態をパターンベクトルとみなし、内積値  $\sum_k S_k S'_k$  で近い  $S_k$  がある環境が同様な反応を返すとする仮定のもとで、あらかじめ与えられた数のカテゴリーを教師なし学習で形成させ、それを SLA に対する状態  $k$  とするものである。

## 5 シミュレーション実験

以上の SLA による自己組織化分散エージェントを、複数タスクを許容するように拡張したジョブショップの動的スケジューリング問題を対象として構成し、計算機シミュレーション実験によりその特性を調べた。対象とするジョブショップは、(1)10種類のジョブ、5種類のタスク、および10台の機械からなり、さらに(2)ジョブとタスクの対応は既知、(3)各機械は複数のタスクを請け負うことが可能だが、タスクそれぞれに対する処理時間は機械ごとに異なるとする。

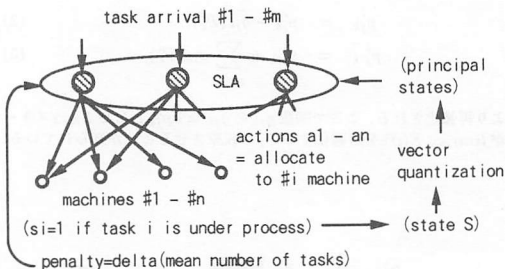


Fig.2 SLA を用いた動的ジョブショップスケジューリングのための分散エージェント構成。

分散エージェントは Fig.2 に示すように、タスクの種類それぞれに対して SLA を配置し、タスクが到着した際に割り当てを行う機械の選択を 10 種類の action とする動作を行うものとする。SLA に対する reward/penalty 値  $\beta$  は、全 SLA に共通に、一つ前の時刻までの 50 時間の移動平均による平均完了タスク数にたいして現在の時刻までの平均完了タスク数の増加、減少をそれぞれ  $\beta = 1, 0$  として与える。学習には  $L_R-I$  スキーム [NARE89] を用いる。具体的には状態  $\phi_j$  に対する action  $\alpha_i$  に対して  $\beta = 0$  のとき、選択確率  $p_{ij}$  を、

$$p_{i^*j} \leftarrow p_{i^*j} + 0.2[1 - p_{i^*j}], \quad (4)$$

$$p_{ij} \leftarrow (1 - 0.2)p_{ij}, \quad i \neq i^*, \quad (5)$$

により更新する。

例題の状態空間は機械の作業中のタスク、ジョブの種類、機械のキューの長さなどの多数の状態変数を含む故に、一般に膨大な次元とならざるを得ない。ここでは実験のため簡単に、各時刻でタスク  $t_i$  がいずれかの機械で処理中か否かを 2 値パターンとした、 $S = \{s_i \in \{1, 0\}\}$  を状態空間とする。これを 4 状態  $S_{SLA}$  にカテゴリー化し、SLA に対する状態入力とする。カテゴリー形成の方法として、ここでは [RUME86] による競合学習則による自己組織化ネットワークを用いる。具体的には状態  $s_j$  に対して

$$\max_i V_i = \sum_j w_{i^*j} s_j, \quad (6)$$

なるユニット  $i$  を発火させ、 $n_S = \sum_j s_j$  として、

$$dw_{i^*j} = 0.05(s_{i^*} / n_S - w_{i^*j}), \quad (7)$$

$$dw_{ij} = 0.0005(s_i / n_S - w_{ij}), \quad i \neq i^* \quad (8)$$

により結合強度を更新させることで、 $S_{SLA}$  が形成される。

以上の分散エージェントを平均仕事到着時間 20、平均作業時間 5 のポアソン到着のもとでシミュレーションし、平均完了タスク数の変化をプロットした結果を Fig.3 に示す。比較のため等確率に機械の選択を行う、学習のないエージェントにより同一のシミュレーションを行った結果を Fig.3 に重ねる。Fig.3 からは等確率のモデルと性能上有意味な差がみられない。ただしこのことは問題設定において機械の性能に大きな偏りをあたえていないことが影響を及ぼしていると考えられる。一方で 4000 時間経過後、ランダム選択で悪い結果が生じる場合にも安定した高いパフォーマンスを達成していることが分かる。機械の性能に対する情報が与えていないにもかかわらず安定した挙動を示すことから、学習が有効に作動していると考えられることができる。

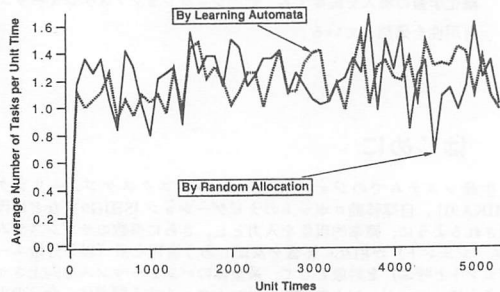


Fig.3 平均完了作業数の時間変化の比較。

## 6 おわりに

確率のマトリックス学習オートマトンを要素として、学習により全体のパフォーマンスを向上させる分散エージェントのモデルを提案した。複数機械割り当てを許したジョブショップの動的スケジューリング問題に適用し、シミュレーションで学習の有効性を確認した。

## 参考文献

- [UEDA91] 上田, 嘉数, “競争モデルによる意志決定モデルに関する研究,” 1991年度精密工学会北海道支部大会講演論文集 (1991).
- [SHIG91] 繁田, 嘉数, “学習オートマトンによる自律移動ロボットのナビゲーションに関する研究,” 1991年度精密工学会北海道支部大会講演論文集 (1991).
- [MIKA91] 三上, 嘉数, “自律分散型生産システムに関する研究,” 1991年度精密工学会秋季大会全国大会講演論文集 (1991).
- [NARE89] Narendra, K.S., and Thathachar, M.A.L., *Learning Automata*, Prentice-Hall (1989).
- [WITT77] Witten, I.H., “An adaptive optimal controller for discrete time Markov environments,” *Inform. and Control*, Vol.34, pp.286-295 (1977).
- [RUME86] Rumerhart, D.E., and Zipser, D., “Feature Discovery by Competitive Learning,” *Parallel Distributed Processing*, The MIT Press, pp.151-193 (1986).