3T－4

行動選択ネットワークによるマルチエージェントの適応学習

長谷川 泰史　米津 光浩　山口 文彦　中西 正和
慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

マルチエージェント (Multi-Agents:MA) 系とは，複数のエージェントで構成される動的な系である．MA系においては状態が複雑であれば複雑であるほど，設計者がシステムの細部まで詳細に設計するトップダウン的な手法よりも，システムにある程度の創発を任せるボトムアップ的な手法が有効である．その際，状態数の増加，行動の創発および適応学習が重要なポイントとなる．

2. 先行研究

MA系の先行研究として追跡問題があげられる．

MA追跡問題

追跡問題とは，複数の捕食エージェントが一定時間内に被食エージェントを捕獲するタスクであり，捕食エージェントには行動選択および学習能力が求められる．Tan[Tan 93] は強化学習を用いて捕食エージェント同士が情報を交換することで，また岩下 [岩下他 95] はやはり強化学習を用いて捕食エージェント同士に情報の交換のない場合でも，それぞれMA追跡問題において強化学習が有効な手法であることを示した．

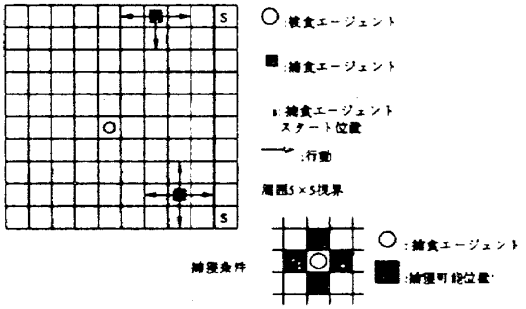

図 1: 追跡問題

強化学習とは，行動の組，状況の組，行動を起こした時に環境から与えられる報酬により定義される．エージェントは報酬の最大化という目的のみを持って，試行錯誤により学習を進める．MA環境における強化学習の問題点の一つは，エージェント数の増加に伴い状態数が爆発的に増加することである．

3. 本研究の手法

本研究では，状態数の削減，ロバスト的な行動選択，環境との相互作用による適応学習という点から行動選択ネットワークに着目する．

3.1 行動選択ネットワーク

行動選択ネットワークとは，行動選択の制御要素あるいは行動要素をノードとし，各ノード間の入出力を通じて行動決定を行なう非常に拡張性に優れたモデルである．代表的なものには ANA(Agent Network Architecture)[Maes 91]，およびその改良 [Maes 92],Behavior Network[Blumberg 96],[鈴木他 97] などがある．本研究ではそのなかでも行動のバランスが良く取れたモデルである [Blumberg 96] のネットワークを用いる．

ノードの構成

各ノードは上位層ノード群からの総和入力，同一層内での排他性入力，下位層からの閾値入力，および目的・環境からの刺激入力により活性値を蓄積させる．ノード内部は，活性値，閾値，実行権，タスク処理から成り，タスク処理の結果としてその出力様式を変化させる．

3.2 本研究の方針

本研究の目的は [Blumberg 96] のネットワークに学習能力をを付加したモデルをMA系に適用することにより，複雑で動的なMA環境において，学習能力を

Action Selection Network in Multi-Agents
YasushiHASEGAWA MitsuhiroYONEZU FumihikoYAMAGUCHI MasakazuNAKANISHI
Department of Computer Science, Faculty of Science and Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223, Japan

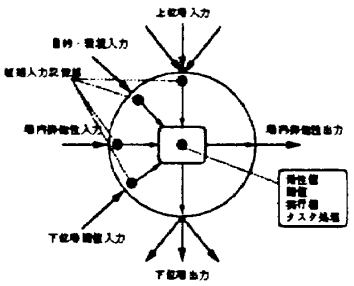

図 2: ノードの構成と入出力関係

有する行動選択ネットワークモデルが有効な手法の一つであることを示すことである. また, エージェント数の増加時において状態数の削減, それによる計算時間の大幅な短縮, および協調タスクにおけるリンクの重みの学習による役割分担の発現を図ることである.

問題設定

グリッドワールドにおける追跡問題への適用を考える. 捕食エージェントは視界内の他のエージェントの相対位置が分かり, 一つ移動するごとに −0.1 の報酬を得るが, そのとき, 被食エージェントとの距離が縮まるならば 0.2 の報酬を得る. また被食エージェントを捕獲すると 20 の報酬を得るものとする. 被食エージェントは単位ステップにおいてランダムに動き回り, かつ捕食エージェントが近付いて来た時には, それから遠ざかる方向へ逃げるものとする.

ネットワーク構成図

各捕食エージェントは, 次のような行動ネットワークで構成されるものとする.

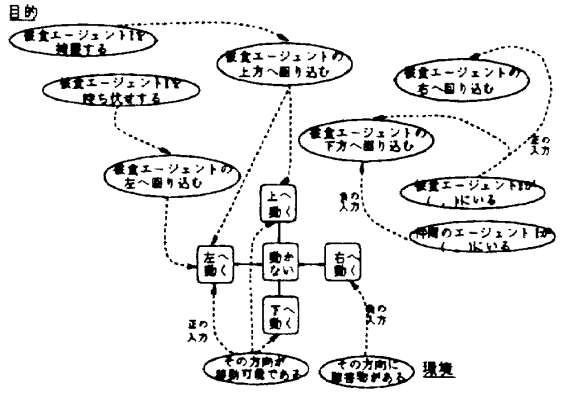

図 3: ネットワーク構成図

下位層ノードにタスク処理として各移動方向を割り当てたノード群を構成する. また, 上位層ノードとして動機づけとなるノード群を割り当てる.

エージェントの学習能力

動的に変化する環境内では, エピソードとよばれる状態遷移系列を単位として学習を行う経験強化型の学習法が環境の変化に適応しやすく, 有効である. そこで, 本研究では経験強化型の学習法である PSP(Profit Sharing Plan) の概念を用い, 環境からの報酬をその獲得行動に関係した全リンクに分配することにより, その重みの強化/減衰を図る.

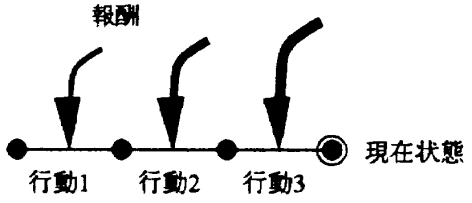

図 4: リンクの強化方法

4. 今後の予定

行動選択ネットワークを MA 追跡問題の個々の捕食エージェントに適用することにより, 状態数の削減を図り行動選択ネットワークが MA 追跡問題における有効な手法の一つであることを確認する.

5. 今後の展望

本研究の応用として, MA 追跡問題におけるエージェント数およびワールドの拡大, さらには MA 系の他の問題への適用が考えられる.

参考文献

[Blumberg 96] Blumberg, B. : *No Bad Dogs: Ethological Lessons for Learning in Hamsterdam, Proc. 4th Int. Conf. on Simulation of Adaptive Behavior, MIT Press, pp. 295–304, 1996*

[Maes 91] P. Maes. : *The Agent Network Architecture(ANA), SIGART Bulletin, Vol. 2, No. 4, pp. 115–120, 1991*

[Maes 92] P. Maes. : *Learning Behavior Networks from Experience, Proceedings of the First European Conference on Artificial Life, pp. 48–57, 1992*

[鈴木他 97] 鈴木他 : 疲労度パラメータを導入した行動選択ネットワークによるエージェントの創発的組織化に関する考察, 人工知能学会誌, *Vol. 12, No. 1, pp. 152-159, 1997*

[Tan 93] Tan. M. : *Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents, Proceedings of the 10th International Conference on Machine Learning, pp. 330–337, 1993*

[岩下他 95] 岩下他 : 強化学習に基づくマルチエージェント系の協調の創発, 第 21 回知能システムシンポジウム予稿集, *pp. 37–42, 1995*

[山村他 95] 山村他 : エージェントの学習, 人工知能学会誌, *Vol. 10, No. 5, pp. 683–689, 1995*