Q-learningを用いたマルチエージェントにおける協調行動獲得に関する研究

Q-learningを用いたマルチエージェントにおける協調行動獲得に関する研究

玉城 斉*　遠藤 聡志**　山田 孝治**

Acquisition of Cooperative Behaviors Using with Q-learning in Multi-agent Environment

Tadashi TAMASHIRO*　Satoshi ENDO*　Koji YAMADA*

Abstract

One of the important issues in intelligent systems and robotics is to develop an efficient method to control multi-agent system. In order to work the multi-agent system well as the problem solver, it's so significant to create cooperative behaviors among the agents. In the multi-agent system, the behaviors of cooperation emerged as the results of suitable role learning by each agent. In this paper, we design some fundamental computer experiments to investigate the emergence of cooperative behaviors in reinforcement learning based multi-agent system. We show that role learning for acquiring cooperative behaviors according to the result of these experiments.

Key Words: multi-agent, cooperative behavior, Q-learning.

1. まえがき

知的システム工学やロボティクスなどの分野において，マルチエージェントのコンセプトが，効果的な問題解決の枠組みとして注目されている．従来の単一エージェントによる問題解決法では，エージェントは，複雑な環境下で現在の状態を認識する高度な状態認識モデルが必要であった．これに対して，マルチエージェントシステムでは，複数のエージェントが個々に局所的な状態を認識し，この情報認識に基づいた協調行動を実現することで，問題の持つ複雑さに対応する．従って，マルチエージェントシステムが効果的に問題を解決するためには，エージェント間での協調行動の獲得が重要となる．マルチエージェントシステムにおける協調行動は，個々のエージェントが解決すべき問題に対する各々の役割を適切に学習することで実現される．本研究では，知的かつ効果的なマルチエージェントシステム設計するために必要な知見を得るための計算機基礎実験を設計し，その結果について議論する．

問題解決のために複数エージェントが協調行動を行なう必要がある問題の一つに追跡問題 [3] がある．追跡問題においては，個々のエージェントは他のエージェントの次行動予測が不能であるため，各々のエージェントが試行錯誤的に役割を学習することで協調を行なう必要がある．

強化学習はエージェントに状態に対する最適行動を学習させる問題である．Q-learningは，典型的な強化学習法の一つで環境から与えられる状態と報酬の組の列を最大化する行動を学習する．そこで，エージェントの学習にはQ-learning [1] を使用する．本報告では，追跡問題を用いた計算機実験を設計し，その結果を示しエージェント系の構成法について議論する．

第一の実験では，追跡問題におけるエージェント間での役割分化の有無の確認を目的とする．

第二および第三の実験では，追跡問題における，役割学習と個々のエージェントの機能差の関係を調べることを目的とする．第二の実験では，エージェントの機能差として，学習率を用いる．Q-learningにおいて，学習率は学習の速度を決定するパラメータである．実験では，学習率の異なるエージェントを用いた場合の，役割学習への影響について調べる．

第三の実験では，エージェントの機能差として，視野を用いる．この実験では，有効視覚範囲の異なるエージェントを用いた場合の，役割学習への影響を調べる．

第一～第三の実験では，目的達成のためには，すべてのエージェントは何らかの役割を担う必要があるという問題の構造をとる．第四の実験では，余剰のエージェントを加えることによりエージェント系全体の振舞いにどの様な影響があるかを調べる．

2. マルチエージェントシステム

マルチエージェントシステムでは，大規模・複雑な問題に対して処理の要素を各エージェントに分担させることで問題解決を行なう．本論文では以下のようにエージェントを定義する．

受理：1995年11月11日
*大学院工学研究科情報工学専攻
(Graduate Student, Information Eng.)
**工学部情報工学科
(Dept. of Information Engineering, Fac. of Eng.)

$$\text{Agent} = \{\mathcal{A}, \mathcal{L}\}$$

ここで，$\mathcal{A}, \mathcal{L}$ は，

$\mathcal{A}$：現在の状態において可能な行動の集合．

$\mathcal{L}$：状態に対応した最適行動を学習する学習機構．

である．エージェント系において，エージェントは環境情報を自己の状態として認知する．そして，その状態に対して可能な行動集合から次の最適行動を過去の学習に基づいて選択する．複雑多機能なエージェントの設計は，システムの即応性やロバスト性を阻害するため，エージェントには通信機能を実装しないことが一般である．本稿では，通信に基づいたエージェント間の協調行動を想定しない．従って，エージェントが互いに各状態毎の最適な行動を学習する必要がある．

本研究では，エージェントの学習法として強化学習を採用する．

3. 強化学習

3.1 強化学習の枠組み

強化学習 (reinforcement learning) は，本来，動物心理学や動物行動学の分野で用いられた用語である．典型的には，報酬 (reward) という特別な入力を手掛かりとして行動パターンを学習する場合に用いられる．広義には，罰による行動の抑制を含め，条件づけといわれる一連の適応現象を実現する学習を指す．

学習者は他の多くの機械学習のように，どの行動を出力すべきかを与えられるのではなく，どの行動が最も高い報酬を得るかを試行錯誤的に探索する．このような単純な原理よって動作する強化学習では，学習対象に対する明示的なモデルを持たずに学習することが可能である．

3.2 マルチエージェントにおける強化学習

マルチエージェントシステムでは，対象問題に対する明示的な知識を持たずに，自律的に動作する処理要素 (エージェント) の相互作用によって問題解決にアプローチするという強化学習との共通点がある．このような枠組みに対して，強化学習は目的や構造において非常に適していると考えられる．

3.3 強化学習の目的

強化学習では，学習エージェントは各時間ステップにおいて観測される状態から行動を決定する．ここで状態とは，学習システムにとっての外部からの入力である．実際にとった行動に対して環境から報酬あるいは罰が与えられるが，報酬の大きさは多くの場合，過去数ステップの行動系列に対して決定される．学習の目的は，ある時間長さにわたる報酬の重み和を最大化することである．報酬と罰を合わせて強化信号 (reinforcement) と呼ぶ．報酬を正の強化 (positive reinforcement)，罰を負の強化 (negative reinforcement) と呼ぶ．

形式的には，時刻 t における強化信号の大きさを r_t とすると，学習者の目的は，現在から未来にわたる強化信号の重み和，式 (1)

$$v_t = \sum_{i=t}^{\infty} \gamma^{i-t} \cdot r_i \tag{1}$$

を最大化することである．ただし，γ は $0 \leq \gamma \leq 1$ の定数であり，割引率 (discount rate) と呼ばれる．$r_t > 0$ の場合には報酬，$r_t < 0$ の場合には罰とする．$\gamma = 0$ の場合は，現在の強化信号のみに着目し未来を無視することになる．逆に $\gamma = 1$ では，どんなに遠い未来でもよいから大きな報酬が得られるほうがよいことになる．すなわち，γ の値の大小によって，どのくらい先の未来までを考慮するかが決まる．しかし，未来の報酬は観測できないので，一般には過去から現在までの強化信号の重み和，式 (2)

$$\hat{v}_t = \sum_{i=0}^{t} \gamma^{t-i} \cdot r_i \tag{2}$$

を v_t の近似として利用する．

4. Q-learning

強化学習法の代表例の一つに，Q-learning がある．Q-leaning では，ルールと呼ばれる状態と行動の組に対する重みを見積もる．この重みを Q 値と呼び，状態と行動の組から重みを導く関数を Q 関数と呼ぶ．

4.1 ルール

ルールとは，エージェントの知覚可能な状態 $x(\in \mathcal{X})$ と選択可能な行動 a の組合せである．状態 x において行動 $a(\in \mathcal{A})$ をとるルールに対する Q 値は，$Q(x,a)$ と記述される．本実験では現在の状態において，Q 値を最大にするルールを選択する方法を用いる．

4.2 Q 値の更新

エージェントが時刻 $t-1$ の状態 x_{t-1} において，行動 a_{t-1} を選択した結果，状態が x_t となり，強化信号 r_t が得られたとすると，Q 値は以下の式 (3) によって更新される．

$$\begin{aligned} Q_t(x_{t-1}, a_{t-1}) = (1-\alpha) Q_{t-1}(x_{t-1}, a_{t-1}) \\ + \alpha(r_t + \gamma \max_{k \in \mathcal{A}} Q_{t-1}(x_t a_k)) \end{aligned} \tag{3}$$

ここで，α は学習率であり，$0 < \alpha \leq 1$ なる定数である．もし，$\alpha = 0$ の場合，Q 値は更新されずエージェントは学習を行なわない．右辺左項は，以前の Q 値がどの程度の割合を占めるのか表す，α の高いエージェントほど，以前の Q 値の占める割合が減少する．右辺右項はこの 時刻 $t-1$ で得られた報酬と現在の状態から見込まれる最大の Q 値であり，α の高いエージェントほど Q 値が大きく更新される．

5. 追跡問題

本研究では，エージェント間の協調が問題解決に必須なモデルとして追跡問題を採用する．追跡問題は，2 次元格子空間上で，複数のハンターエージェントが，ランダムに行動する獲物エージェントを捕獲することを目的としたマルチエージェント系のモデルである．

実験の設計のために以下のように問題パラメータを設定した．

環境は $n \times n$ の 2 次元トーラス構造とする．(Fig. 1)

この環境におけるエージェントの行動は以下に示す通りである．各エージェントは各タイムステップ毎に 1 マス移動する，あるいは止まるといった行動を行なう．また，エージェントは同じマス上に同時に存在できる．

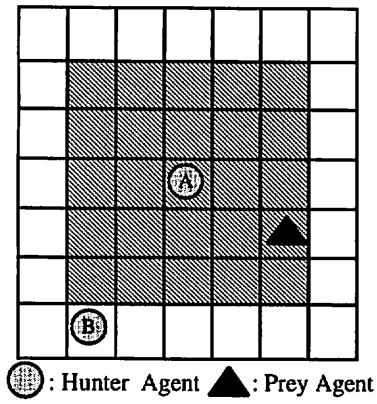

Fig. 1. 追跡問題

捕獲条件は，2つのハンターエージェントが獲物エージェントを両側から挟むこととする．(Fig. 2)

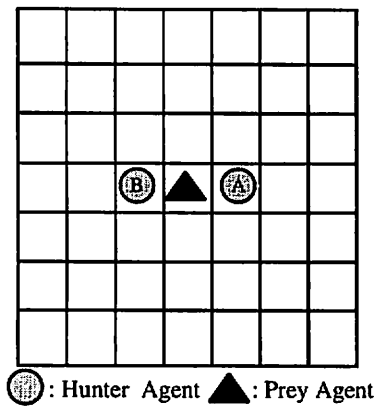

Fig. 2. 捕獲条件

ハンターエージェントの学習機構としてQ-learningを採用する．ハンターエージェントには，視覚としてエージェント自身からの他のエージェントの相対位置が与えられる．また，エージェントの視野は限られている場合があり，視野の深さがdとするとハンターエージェントは自分を中心に$(2d+1)\times(2d+1)$マスの中の他のエージェントを知覚できる．例えば，$d=2$のときエージェントの視野は5×5マスで，fig. 1においてエージェントAが受けとる状態は，$(-1,2)$でエージェントBは視野にいないので相対位置は与えられない．また，ハンターと獲物の区別が可能である．ここで，獲物を捕らえることによるrewardは1.0とする．

ハンターエージェントの初期配置は，ランダムで常に獲物とは一定距離以上離れた位置に配置される．獲物捕獲までを一試行とし，捕獲後に再配置し試行を繰り返しエージェントの学習を行なう．

以上の環境で実験を設計する．

6. 計算機実験

6.1 実験1：役割学習の確認

第一の実験では，追跡問題におけるハンターエージェント間での役割分化の有無の確認を目的とする．

ここで，環境の大きさは11×11とする．ハンターエージェントおよび獲物エージェントの数は，それぞれ2，1とする．ハンターエージェントの能力パラメータとして．この実験では学習率，視野の深さを用い以下のように値を設定する．学習率は，$\alpha=0.2,0.4,0.6,0.8$の各々について実験し，視野の深さ$d=5$に固定する．ここで視野の深さ$d=5$とはハンターエージェントが環境全体を知覚できることをあらわす．

6.2 実験2：機能差の実験(学習率)

この実験ではエージェントに機能差を与えることが役割学習にどのような影響を及ぼすかを調査する．一般に，エージェント系を性質の異なる異種の系で構成することが，エージェント系の全体の性能向上につながるということが指摘されている[4]．実験では，機能差にQ-learningの主要なパラメータである学習率を用い，異なる学習率でエージェント系を構成した場合の役割学習の有無を調査する．

エージェントの学習率の組合せα_A,α_Bは，以下のように設定する．

- $\alpha_A=0.2$と$\alpha_B=0.4$のエージェント系
  低学習率同士の異種エージェント系
- $\alpha_A=0.2$と$\alpha_B=0.8$のエージェント系
  低学習率及び高学習率の異種エージェント系
- $\alpha_A=\alpha_B=0.4$のエージェント系
  同種エージェント系(比較対象：実験1に同じ)
- $\alpha_A=0.4$と$\alpha_B=0.8$のエージェント系
  中学習率及び高学習率の異種エージェント系
- $\alpha_A=0.6$と$\alpha_B=0.8$のエージェント系
  高学習率同士の異種エージェント系

以上について，獲物捕獲までのステップ数に関する比較を同種エージェント系と行なうために，10回の繰り返し実験を行ないその結果についても考察を行なう．

6.3 実験3：機能差の実験(視野)

第三の実験では，エージェントの機能差として，視野を用いる．実験の目的は，有効視覚範囲の異なるエージェントを用いた場合の，役割学習への影響を調査することである．これまでの実験では，エージェントの視野は環境全体と同一であった．ここでは，エージェントの視野を制限し，環境の大きさを20×20とする．エージェントの視野の深さは，それぞれ，$d_A=2,d_B=4$とする．比較の対象として視野の深さを同一$(d_A=d_B=3)$とした場合のエージェント系の実験を行なう．

6.4 実験4：余剰エージェントの実験

この実験では，余剰のエージェントを加えた場合のシステムの振る舞いを調査することを目的とする．これまでの実験では，すべてのハンターエージェントが目的達成のために，何らかの役割を担う必要があった．ここでは，ハンターエージェントの数を3とし，すなわち，余剰エージェントをエージェント系に追加することでそれまでの実験と比較する．

本来，マルチエージェントの扱う問題環境下においては，目的に対する適切なエージェント数は不明であり，エージェントの数は目的に対して十分な数が与えられることが一般である．そこで，目的の達成に不要なエージェントを加え

た場合の系の振る舞いを検証する．

7. 実験結果及び考察

7.1 実験1 - 役割学習

fig. 3は，$\alpha = 0.2, 0.4, 0.6, 0.8$ の4つの場合のエージェント系において，エージェントが獲物を捕らえるまでの1,000試行毎の平均ステップ数の推移を示すものである．

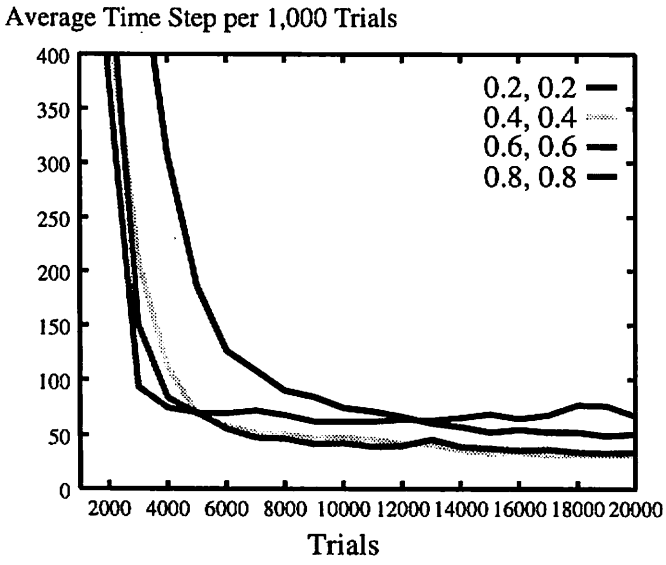

Fig. 3. 学習率同種の系における獲物捕獲のステップ数の変化

ここで，縦軸は平均ステップ数，横軸は試行数を示す．fig. 3は，すべてのエージェント系においてエージェントは2,000試行までは，獲物を効果的に捕らえることができず，獲物捕獲平均ステップ数は，150ステップ以上要している．しかし，8,000試行を過ぎるとエージェントは獲物を効果的に捕らえるようになりエージェントは平均100ステップ以下で獲物を捕獲できるようになる．また，学習率0.4と0.6のエージェント系において，8,000試行以降では，平均50以下で獲物を捕獲している．獲物捕獲までのステップ数の比較から，学習率0.4と0.6エージェント系が効果的に問題解決を行なっている系であることが分かる．

次に，エージェントAの捕獲位置の割合の推移を示したグラフをfig. 4 ～ 7に示す．グラフの縦軸は獲物捕獲位置の割合を，横軸は試行回数を示す．

fig. 4は，$\alpha = 0.2$ のエージェント系での獲物捕獲位置の割合の推移を示している．エージェントの捕獲位置は7,000から12,000試行以外ではほとんど違いが見られない．$\alpha = 0.2$ の系は，獲物捕獲までのステップ数が減少していない系であり，このような系においては明確な役割分化が見られない．

fig. 5は，$\alpha = 0.4$ のエージェント系での結果である．図より4,000試行までは，行動差が見られないが，6,000試行以降では，捕獲位置が右と上に大きく偏る．fig. 3の6,000試行付近は，十分に学習が進んだ状況を示しており役割分化とエージェントの機能向上は役割と密接な関わりがある．

また，fig. 6は，$\alpha = 0.6$ のエージェント系の結果である．3,000試行付近から，右から獲物を捕獲するという行動獲得が始まり，他のエージェント系と比較して最も早い役割分化が現れている．また，5,000試行以降で役割分化は安定し80％以上の割合で右から捕獲を行なっている．fig. 3の3,000試行以降に，すでに効果的に獲物捕獲を行なっており役割分化とエージェントの機能向上が密接に関わっていることが分かる．

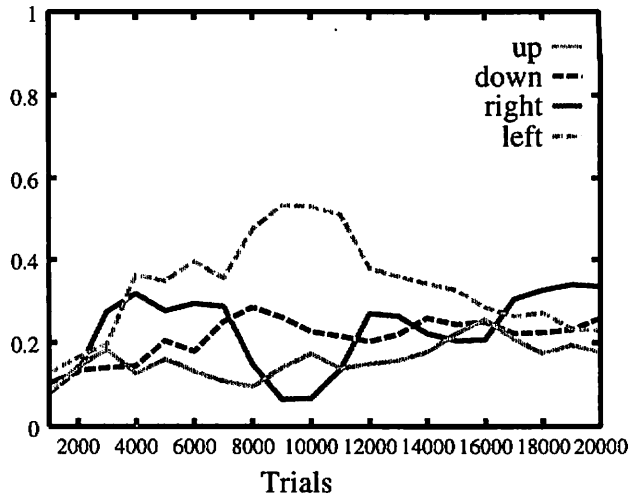

Fig. 4. 学習率0.2 同種の系におけるエージェントAの捕獲位置の推移

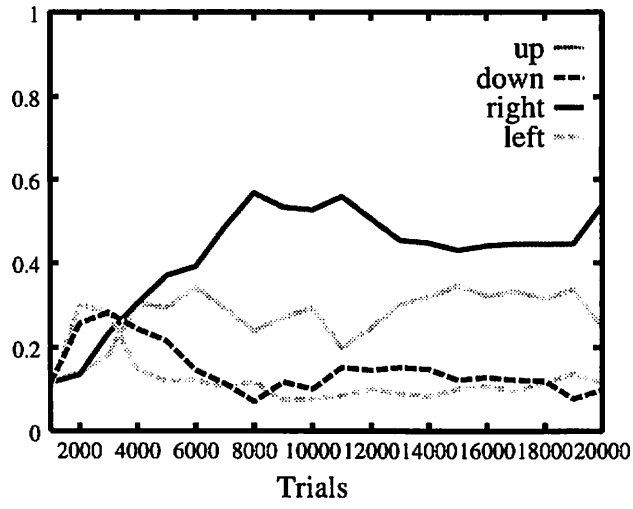

Fig. 5. 学習率0.4 同種の系におけるエージェントAの捕獲位置の推移

fig. 7は，学習率が0.8のエージェント系の結果である．ハンターエージェントAは獲物を上，下または左から捕らえているがハンターエージェントの役割が安定せず，20,000試行を終えた時点では，主に左から獲物を捕らえるように役割の逆転がみられる．また，fig. 3を見ると4,000試行以降からステップ数が減少しなくなり，エージェントの学習に行き詰まりが見られる．これは，エージェントの役割が安定しないことに原因すると考えられる．

以下に$\alpha = 0.4$ のエージェント系においてエージェントAが最終的に獲得した役割を示す．fig. 8は，20,000試行まで実験を行なった$\alpha = 0.4$ のエージェント系での19,000から20,000試行におけるエージェントAとエージェントBの獲物捕獲位置の割合である．右上から，時計回りに上，

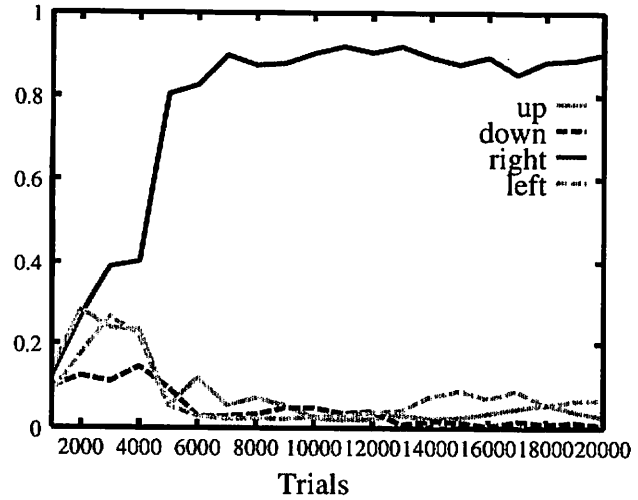

Fig. 6. 学習率 0.6 同種の系におけるエージェント A の捕獲位置の推移

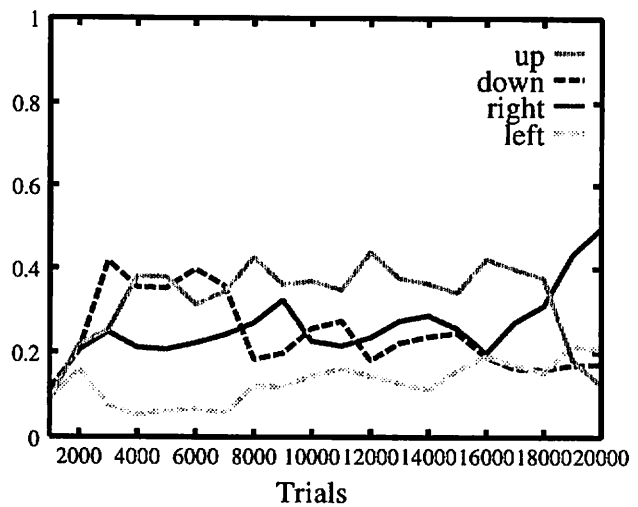

Fig. 7. 学習率 0.8 同種の系におけるエージェント A の捕獲位置の推移

下，右，左の捕獲位置の割合である．fig. 8より，学習の結果エージェント A の獲物捕獲位置は 53.8 %の割合で右，24.8 %の割合で上の位置を占めている．また，同様にエージェント B の獲物捕獲位置は左，下の順に高い割合を占めている．

これは，エージェント A，B がそれぞれに右，上と左，下から獲物捕獲を行なうという異なる行動を学習したことを示している．これらの行動は各エージェントの役割として見ることが出来る．

以上の結果から，$\alpha = 0.4, 0.6$ のように役割を学習しているエージェント系は，獲物を効率的に捕獲していることが分かる．このことからエージェントに協調行動を学習させるためにはエージェントの役割学習が重要であると考えられる．

7.2 実験 2 - 機能差 (学習率)

fig. 9では，エージェント系の獲物捕獲までのステップ数の推移を示している．縦軸に 1,000 試行毎の平均ステップ数を，横軸には試行回数を示している．

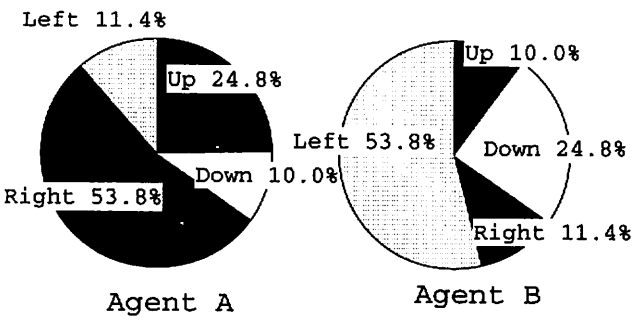

Fig. 8. エージェントの獲物捕獲位置の割合

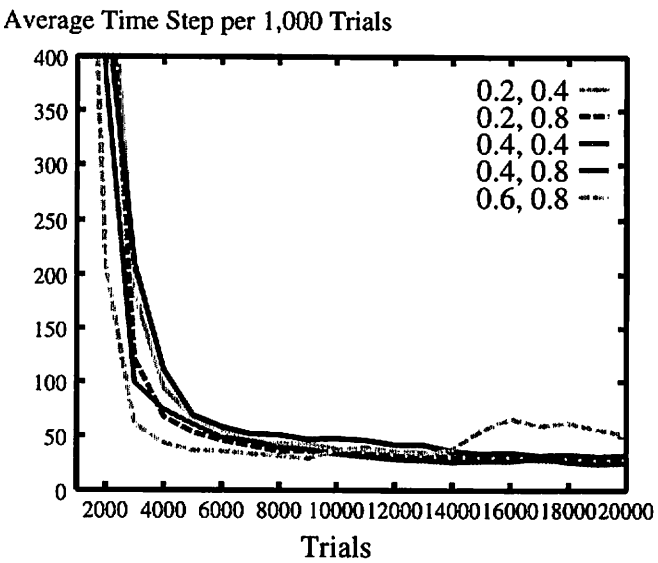

Fig. 9. 学習率異種の系における獲物捕獲までのステップ数

図より，全てのエージェント系は 5,000 試行までに 100 ステップを下回っていることが分かり，fig. 3と比較しても学習の速度が速いことが分かる．また，$\alpha_A = 0.6, \alpha_B = 0.8$ の系を除き，どの系も最終的に平均 50 ステップ以下で獲物を捕獲している．$\alpha_A = 0.6, \alpha_B = 0.8$ の系では，3,000 試行以降で，平均ステップ数が 100 ステップ以下になり，他の異種エージェント系に比べても学習の速度が速いが，14,000 試行以降ステップ数が増加し過学習を起こしている．

以下 fig. 10 ～ 13にエージェントの獲物捕獲位置の割合の推移を示す．縦軸に 1,000 試行毎の獲物捕獲位置の割合，横軸に試行回数を示す．

fig. 10に示すように，$\alpha_A = 0.2, \alpha_B = 0.4$ のエージェント系において，エージェント A は主に上と右から獲物を捕獲している．エージェント B は逆に下と左から獲物を捕らえるようになっており，それぞれのエージェントが役割学習を行なっていることが分かる．また，実験 1 において学習率 0.2, 0.4 同種のエージェント系 (fig. 4, 5) と比較してその学習率の組合せである $\alpha_A = 0.2, \alpha_B = 0.4$ の系は，初期の試行においてはほとんど異差がない．最終的には 20,000 試行において上と右から獲物を捕獲するという役割を学習している．また，ステップ数の比較では，fig. 3より，同種の系の $\alpha = 0.4$ とほぼ等価であることから，学習率を下げ

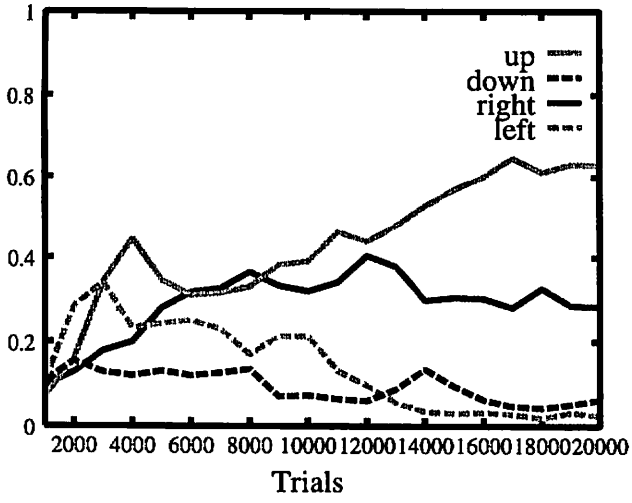

Fig. 10. 学習率 0.2, 0.4 の系におけるエージェント A の捕獲位置の推移

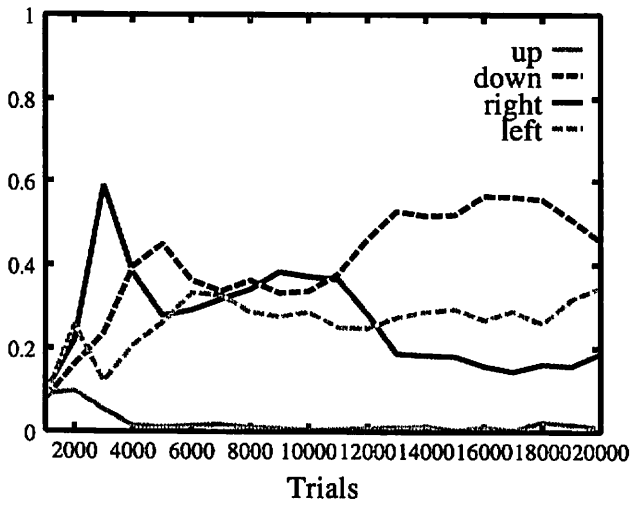

Fig. 11. 学習率 0.2, 0.8 の系におけるエージェント A の捕獲位置の推移

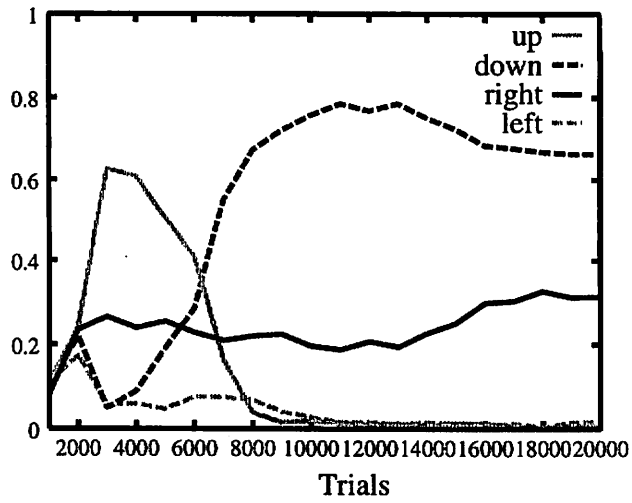

Fig. 12. 学習率 0.4, 0.8 の系におけるエージェント A の捕獲位置の推移

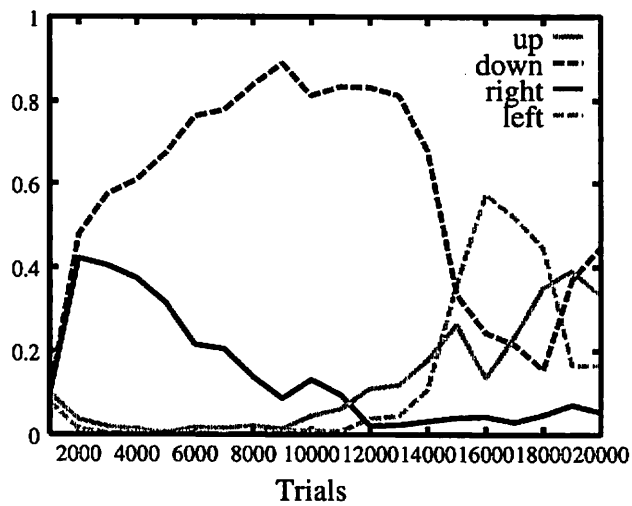

Fig. 13. 学習率 0.6, 0.8 の系におけるエージェント A の捕獲位置の推移

た異種エージェント系の機能低下は見られなかった.

fig. 11では, $\alpha_A = 0.2, \alpha_B = 0.8$ のエージェント系のエージェントの捕獲位置を示している. この系と $\alpha_A = \alpha_B = 0.2, \alpha_A = \alpha_B = 0.8$ の同種の系とを比較すると. fig. 10の場合と同様に同種のエージェント系と異なり最終的に役割学習を行なっていることが分かる.

fig. 12は, $\alpha_A = 0.4, \alpha_B = 0.8$ のエージェント系の獲物捕獲位置の推移を示している. エージェント A は, 初期の3,000試行にはすでに異なる行動を学習し, 役割を獲得している. しかし, 6,000試行目に行動が大きく変化し, 役割は異なるものとなっている. $\alpha_A = \alpha_B = 0.4, \alpha_A = \alpha_B = 0.8$ の同種の系と比較を行なうと, 役割の学習が速いことが分かる. 最終的に獲得された役割は安定している.

fig. 13は, $\alpha_A = 0.6, \alpha_B = 0.8$ のエージェント系の獲物捕獲位置の推移である. エージェント A は, 2,000試行付近ですでにエージェント B と異なった行動を学習している. fig. 9を見ると, この系のステップ数の減少は, 他の系と比較して早いことが分かる. しかし, 14,000試行から学習した行動に変化が起き役割の変化と共にステップ数も増加したいる.

以上の結果から,

(1) 役割学習が進んだ系では, 学習が早く進みステップ数が減少すること.

(2) 学習した役割が大きく変化するとステップ数が増加し, エージェントの協調が難しくなること.

が分かる.

また, 学習率 0.4 と他の学習率との組合せ及び, 学習率 0.6, 0.8 の組合せ, 各々のエージェント系の獲物捕獲までのステップ数の減少を fig. 14に示す.

fig. 14より, 学習率の異なるエージェント系は, 先の fig. 3の結果において性能の良かった学習率 0.4 同士のエージェント系とほぼ同等の性能を示している. これは学習率の異なるエージェント系においてエージェントの能力の差異によってエージェントが役割を早くかつ安定的に学習していることを示していると考えられる.

以上の結果から, 学習率の異差がエージェントの役割学習に効果的に働き結果として, マルチエージェント系の性能を向上させていることが分かる.

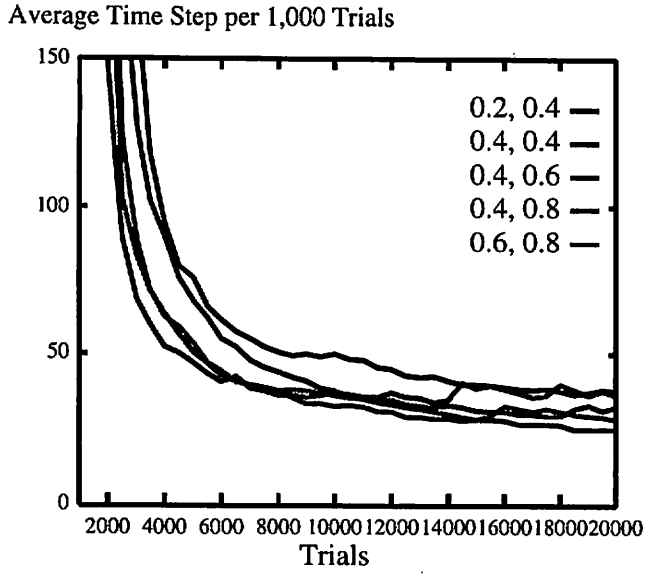

Fig. 14. 学習率異種の系における獲物捕獲までのステップ数の変化 (10回の平均)

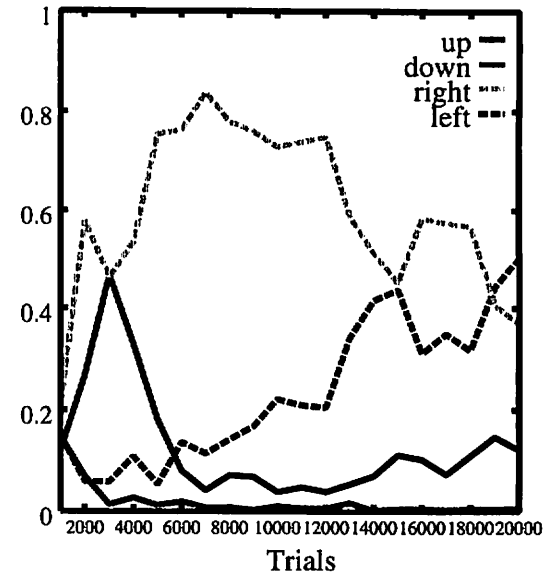

Fig. 15. $d_A = d_B = 3$ のエージェント系の獲物捕獲位置の推移

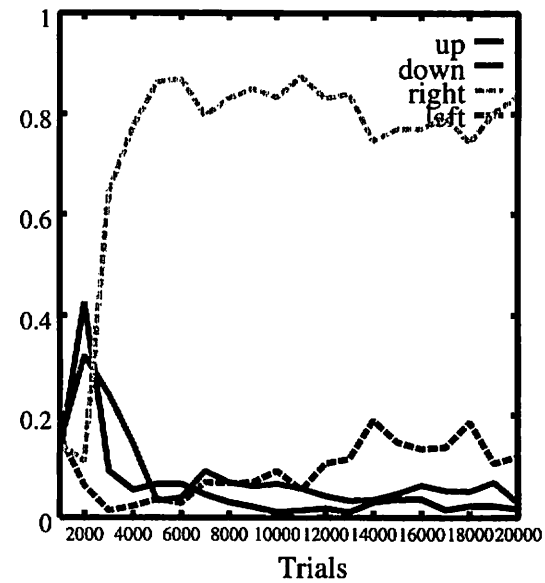

Fig. 16. $d_A = 2, d_B = 4$ のエージェント系の獲物捕獲位置の推移

実験結果より，エージェント間の協調行動の獲得にはエージェントの役割学習が重要であることが確認された．また，学習率の異なったエージェント系を構成することによりエージェントの役割学習が容易になり，マルチエージェント系の性能向上に有効であることが確認された．

7.3 実験 3 - 機能差 (視野)

結果を fig. 15,16に示す．図は，各 1,000 試行毎のハンターエージェント A の獲物捕獲位置の割合を示している．縦軸に捕獲位置の割合，横軸には試行回数である．fig. 15 は，全てのハンターエージェントの視野の深さを 3 としたエージェント系での捕獲位置の割合を示している．ハンターエージェントは 5,000 試行以降主に右から獲物を捕獲するという行動を学習しているが，14,000 試行以降，左からの獲物捕獲という行動をとるようになる．エージェント数は 2 であるから，エージェント B も同様に左と右から獲物を捕らえるという行動を学習している．従って，両エージェント間の行動に差が見られなくなり，役割が安定しない．fig. 16は視野の深さが異なるエージェント系での捕獲位置の割合の推移を表している．図より，ハンターエージェント A は 4,000 試行以降，獲物を 70 %以上の割合で右から捕獲しており，行動も安定している．このことから，明確な役割学習を行なっていることが分かる．

以上の結果から，エージェントの機能差として視野を用いた場合，役割学習が明確に起こり役割分化の安定性が向上する．

7.4 実験 4 - 余剰エージェント

ハンターエージェント数を 3 とし，余剰エージェントを加えた場合の実験結果を fig. 17 ～ 20に示す．fig. 17は，3 エージェントの獲物捕獲位置の割合を示している．fig. 18, 19, 20は，それらのエージェントの捕獲位置の推移を示している．fig. 17より，エージェント C は獲物を 42.6 %の割合で左から捕獲している．一方，エージェント A,B は，獲物を右と左から捕らえていて明確な行動の違いが見られない．従って，エージェント C は役割学習を行なうが，他のエージェントには役割学習が見られない．その理由としてエージェント A,B が C との協調によって学習が進んだ場合，A と B の協調による獲物捕獲結果の学習は以前の役割学習の結果に悪影響を及ぼす．例えば，エージェント A がエージェント C と協調するために右からの獲物捕獲行動の割合が増えたとすると必然的に左から獲物捕獲を行なわなくなり，エージェント B と獲物を捕らえることが出来なくなる．fig. 20からも，エージェント C は安定した役割を学習していることが確認できるが，fig. 18, 19 から分かるようにエージェント A,B は明確な役割の違いが見られない．

従って，余剰なエージェントを加えた場合，2 エージェントの行動が類似のものとなり結果として，エージェント間の安定的な役割学習が行なわれないことが確認された．

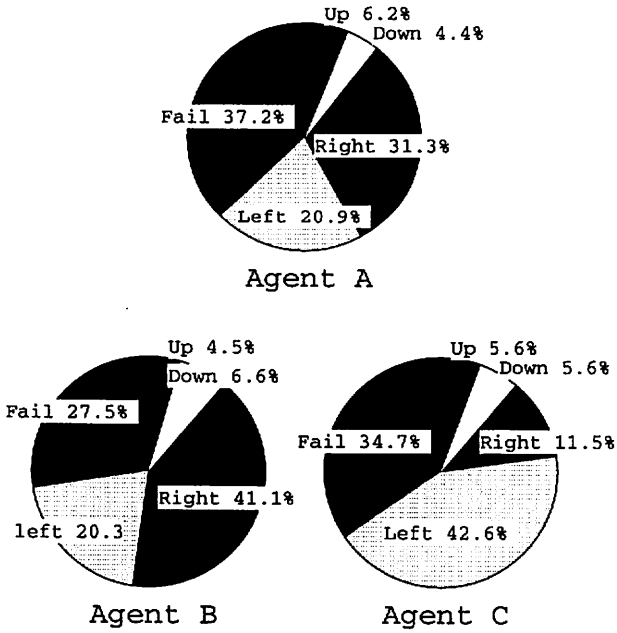

Fig. 17. 3つのハンターエージェントの役割

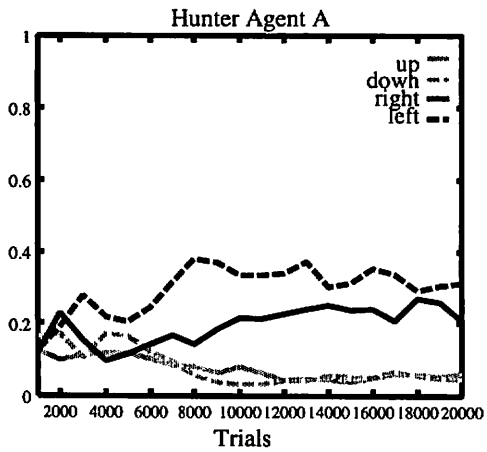

Fig. 18. 余剰エージェントを加えた場合でのエージェント A の獲物捕獲位置の推移

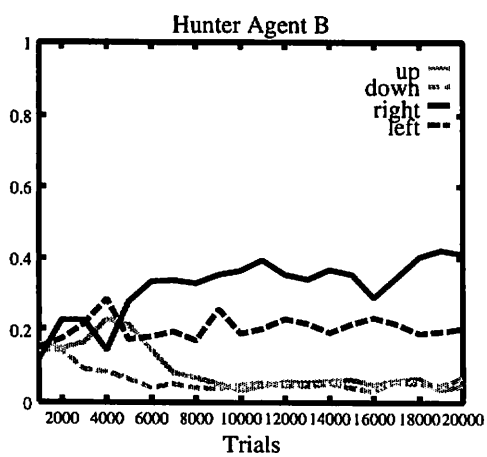

Fig. 19. 余剰エージェントを加えた場合でのエージェント B の獲物捕獲位置の推移

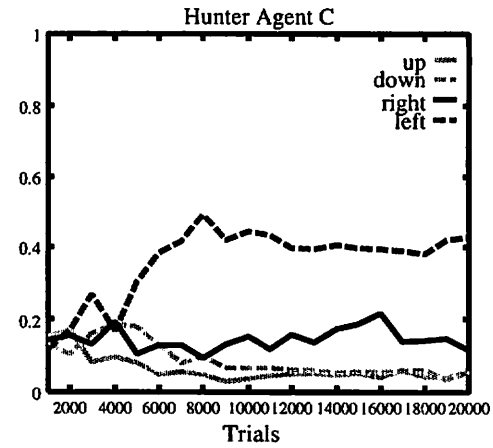

Fig. 20. 余剰エージェントを加えた場合でのエージェント C の獲物捕獲位置の推移

8. おわりに

本報告では，マルチエージェントによる協調行動獲得のため，役割学習に関する基礎計算機実験を行なった．

その結果，以下の結論を得た．

- 役割学習が明確に現れたエージェント系では，系全体の能力向上が見られた．
- 学習率の異なる系を構成することによって，各エージェントの行動の違いが明確に現れ，役割学習が容易になる．
- 余剰なエージェントを加えた場合，役割学習に悪影響を及ぼす．結果としてエージェントの追加による期待される系の能力向上が見られない．

文 献

[1] Watkins,C.J.C.H. and Dayan,P. Technical Note:Q-Learning, *Machine Learning*, Vol.8, No.3, pp.279-292, 1992.

[2] Tan,M Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents, *Proceedings of the Tenth International Conference on Machine Learning*, pp330-337,Morgan Kaufmann, 1993.

[3] Benda,M., Jagannathan,V. and Dodhialla,R. On Optimal Cooperation of Knowledgen Sources, *Technical Report*, BCS-G201-28,Boeing AI Center, 1985.

[4] 河石 勇 山田 誠二 豊田 順一 異種学習エージェント系における経験の共有と学習効率, MACC'95 Online Proceedings 研究会資料シリーズ No.1, 1995.

[5] 畝見 達夫 強化学習 Reinforcement Learning, 人工知能学会誌 Vol.9 No.6 Nov. 1994, 1994.

[6] 寺邊 正大 椹木 哲夫 片井 修 鷲尾 隆 マルチエージェント環境下での情報の共有と組織学習, MACC'95 Online Proceedings 研究会資料シリーズ No.1, 1995.

[7] 山村 雅幸 エージェントの学習, 人工知能学会誌 Vol.10 No.5 1995 9, 1995.

[8] 石田 亨 エージェントを考える, 人工知能学会誌 Vol.10 No.5 1995 9, 1995.