記憶の階層構造のエージェントへの実装 と 協調の創発

5AF－6

山崎和子*
東京情報大学
e-mail:yamasaki@rsch.tuis.ac.jp

1 はじめに

強化学習の研究では、環境からの知覚をそのまま学習器への入力に用いることが多い。しかし、エージェントが多数存在するシステムではこの場合の数が膨大になり、学習が困難になる。なんらかの一般化が必要であるが、これを設計者がアプリオリに行う研究が多い。環境の変化にも耐えうる柔軟なシステムの作成のためには、これをエージェント自らがやるべきにもかかわらず、その試みは少ない。[4]そこで、記憶に階層構造をつくり、学習の進捗や環境の変化にあわせて動的に再構築させるしくみを「知覚」から「学習器への入力」の間にいれ、知覚の一般化をさせた。

2 知覚の一般化

人間は、知覚の特徴を目的にあわせて一般化する。また、知覚の変化は無意識のうちに潜在意識の中に蓄積され、その変化が閾値超えた時、意識の中でその変化が認識されると考えられる。それに沿って、次のようなモデルを考えた.

1. 知覚そのものを一般化するのではなく、知覚及びその時点で学習した結果も含めて一般化する。

2. エージェントの記憶の中に、潜在意識(第1層目),意識(第2層目),全体としての認識(第3層目)という階層構造をもたせる。

1層目では,知覚はそのままFIFOで記憶さる。2層目では,1層目の記憶が,定期的に,知覚及び学習結果に基づき分類,保存される。3層目でには,環境の全体としての記憶をさせた。

3 シミュレーション

エージェント間の「競合」が存在し、「協調」が生じる可能性のある問題として次のようなシミュレーションを行った。任意の大きさの格子に図1のような向きを持った格子からなる通路がある。(x、y方向の境界はサイクリックにつながっている) ここに、A、B2種類のエージェントがそれぞれ10個づつ、ランダムに配置した。A（B）はこの通路をx座標の正(負)方向に1周すると、報酬が与えられる。エージェントは格子の向きを観測できるので、エージェントの数が少ない時には学習することは簡単である。エージェントの数が多くなると、異種のエージェントとぶつかりあいながら通路を進む.。この時エージェントが学習によってどのように競合を回避しながら協調行動をとるかを観測したい。学習方法は,いわゆる経験強化型[2]の中から,まず[1][3]Profit Sharingを用いた。1層目から2層目の一般化は,知覚ではなしに学習した遷移確率について,クラスター分析を行った.以下の4種類のシステムについてシミュレーションを行った

（S1）3層の記憶を持ち、10000ステップごとにクラスター分析を行うシステム

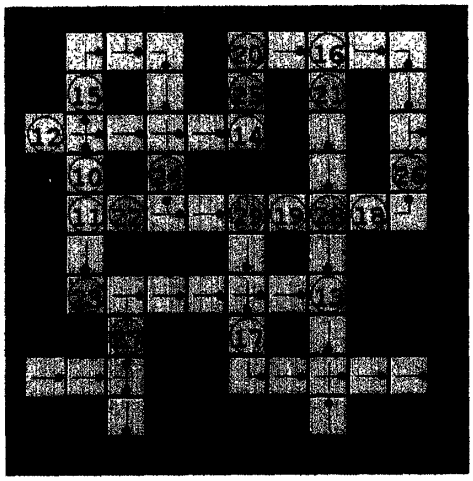
（S2）3層の記憶を持ち、10000ステップ経過時に1回のみクラスター分析を行うシステム

（S3）上部1層の知覚を統一した記憶のみもつシステム

（S4）下部1層の知覚をＦＩＦＯで記憶したもののみ持つもの

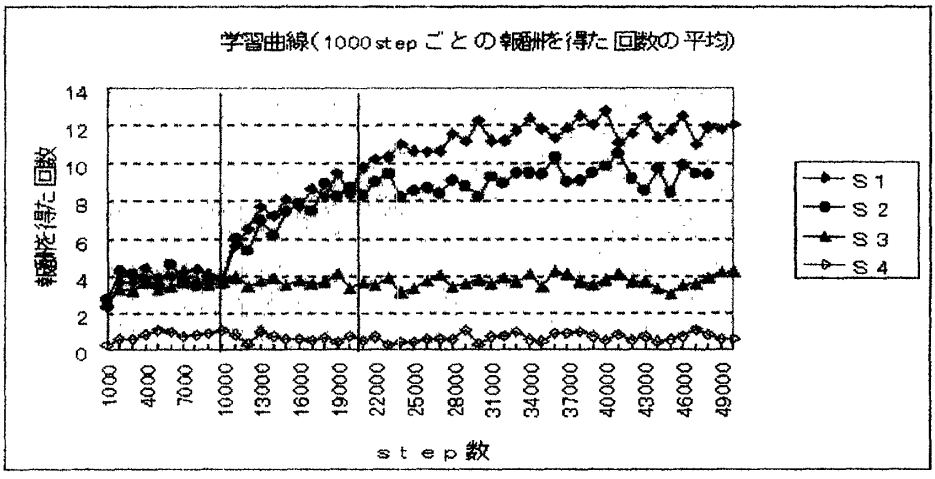
*Agent with layerd memory and emergent cooperation Kazuko Yamasaki Tokyo University of Information Science 1200-2 Tanitou,Wakaba,Chiba 265,Japan

下部の１層の知覚の記憶の容量は大きさを３００とした。１回を５００００ステップとし、それぞれ５回、１００個のエージェントについて平均した。



4　結果

図2はそれぞれの場合についての学習曲線をである。（Ｓ１）（Ｓ２）（Ｓ３）のシステムは10000ステップまではその方法に差はない。10000ステップでａ，ｂのシステムはクラスター分析を行い中間の2層目を生成する。20000ステップまでは（Ｓ１）（Ｓ２）のシステムはその方法に差はない。20000ステップの時に（Ｓ１）のシステムは中間の2層目を再構築する。以上の結果より、記憶の階層構造、その再構築が有効に働いていることがわかった。各エージェントについて、同種又は異種のエージェントが隣り合う知覚が観測される回数を調べた。あるエージェントにとって異種のエージェント数は１０個、同種のエージェントの数は９個であるから、全くランダムの場合には同種のエージェントが出会う回数／異種のエージェントが出会う回数の比は０．９になる。シミュレーションによると、エージェントの報酬が得られる回転方向のパスに隣り合う割合は同種／異種＝１．９５それと逆方向のパスに隣り合う割合は同種／異種＝１．０３であった。いずれもランダムな場合の値０．９より明らかに大きい。これは、同種のエージェントが群れを作って行動し、少しでも、異種のエージェントとの摩擦を減らそうとした結果だと思われる。目でシミュレーションの画面を観察した場合、境界のはっきりとした群れではないが、同種のエージェントが集まっている時間が長く観察できた。



参考文献

[1] J.J.Grefenstette,Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms,Machine Learning 3,pp.225-245,(1988)

[2] 山村雅之、宮崎和光、小林重信、エージェントの学習、人工知能学会誌、Vol.10,No.5,pp.683-689,(1995)

[3] 宮崎和光、山村雅之、小林重信、強化学習における報酬割当の理論的考察、人工知能学会誌、Vol.9,No.4,pp.580-587,(1994)

[4] 上野敦志、堀浩一、中須賀真一、報酬に基づく状況認識と状況に基づく行動選択の同時学習、人工知能学会研究会資料、SIG-FAI-9503-6(03/01),pp.38-45,(1996)