1B3-4

学習結果のカテゴリ化によるマルチタスクへの適応

Adaptation to multitask by categorization of self-memory

大村英史[*1] 片上大輔[*1] 新田克己[*1]
Hidefumi Ohmura Daisuke Katagami Katsumi Nitta

[*1] 東京工業大学 大学院総合理工学研究科
Tokyo Institute of Technology Interdisciplinary Graduate School of Science and Engineering

In this paper, we propose a learning method for an agent to interact with various types effectively. This method adopts *Q*-learning and a clustering algorithm: ward method. This method runs like categorization theory of human for agents. We have experimented with this method. The agent categorizes self-memory of mazes, and run to new a maze. We observe and discuss behavior of agents.

1. はじめに

近年，ロボットのようなエージェントが様々な形で人間の実生活に関わりつつある．例えば，人間の生活を補助する生活支援ロボットや介護ロボット，趣味･娯楽を目的としたエンターテイメントロボット，情報ネットワークとの情報交換を行うメディアエージェントなどが挙げられる．人間はロボットやコンピュータなどの機械を生物のように扱うと言われている[Reeves 96]．現在のエージェントは，音声会話，ジェスチャー，表情などといったインターフェースがあるものの，人間同士のインタラクションとは程遠い．このためエージェントとインタラクションを行う人間は，人間とのインタラクションに比べ負担が多いと考えられる．これを解決するためには，人間同士のようなことができるエージェントの枠組みが必要である．

人とエージェントとの間の自然なインタラクションを目指す研究はヒューマンエージェントインタラクション(HAI)[山田 02]と呼ばれており，活発に行われている．また，このような人間に近い能力を「社会的知能」「社会的知性」などと呼ばれており，社会の中で生活するための必要な能力として，エージェントへの適用が試みられている．このような能力として，我々はエージェントが接する相手の類似性を見つけ出し，直接経験していないことでも頭の中で経験するといった人間のような能力を付加させるシステムの開発を行った[片上 05]．また，[片桐 03][中島 04]では相手との関わりにおいて発生する社会的な感情を心のモデルとし，心のモデルとパーソナリティをエージェントに付加することで社会性知性のひとつとして社会的応答を検討している．

[片上 05]で提案した MULA はエージェントが類似性を用いることによって，経験したことのないことを類似した他人と経験していればそれを自分の物として利用する．この類似性をはかる方法は，距離計算を用いており一対一の比較である．本研究では複数との比較をあらかじめ作っておき，使用したいときに類似経験を用いることができるシステムを提案する．これを実現するために，人間の能力のひとつである「カテゴリ化」に注目する．カテゴリ化は古くから社会心理学の分野で研究されており，重要な人間の能力だと考えられている．この能力はデータマイニングなどに応用されており，様々な手法が提案されている．この機能を，学習機能をもったエージェントに付加させて行う．

第 2 章では本研究で用いるカテゴリ化と学習について説明をする．第 3 章では社会性を目論んだエージェントのための新しい学習手法を提案する．第 4 章で実験により，システムを動かし，システムの検討をする．

連絡先：大村英史，東京工業大学総合理工学研究科，知能システム科学専攻
ohmura@ntt.dis.titech.ac.jp

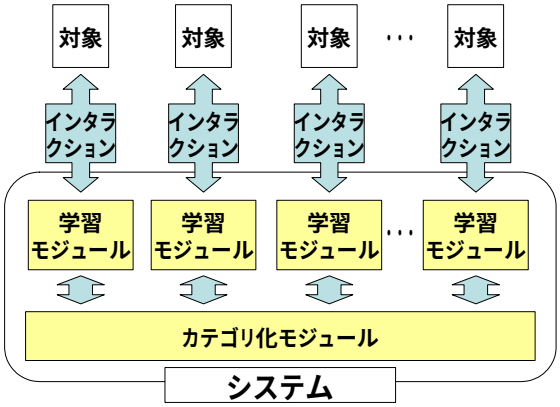

図 1　提案システム概要

2. カテゴリ化と学習

2.1 カテゴリ化

一般的に人間は，多くの情報をある意味のまとまりに分類して頭の中に蓄積していると考えられている．これをカテゴリ化と呼んでいる．カテゴリ化は人間の認知能力に影響を与える物である．例えば，ステレオタイプと呼ばれているものがある．これは偏見や差別の基本になっているといわれることがあるが，知覚者において認知することが単純で効率が良いものにしている．これもカテゴリ化である．

このカテゴリ化の能力は情報処理の分野ではデータマイニングの手法，クラスタリングに応用されている．本研究ではこのクラスタリングアルゴリズムを用いて類似経験をリアルタイムに獲得する方法を実現する．

2.2 学習

本研究では強化学習を用いる．強化学習は未知の環境に対して試行錯誤を行い，行為を学習する手法である．しかしながら，大規模な学習をする際，状態空間の広がりにより膨大な時間がかかってしまう．そこで，学習器を複数もたせ，階層構造や並列構造をした学習システムが提案されている[山岸 03] [鮫島 02]．

本研究では MULA と同様に複数の学習器をあらかじめ用意してある．複数を持たせることで，クラスタリング手法を利用し類似経験の獲得をできるようにする．

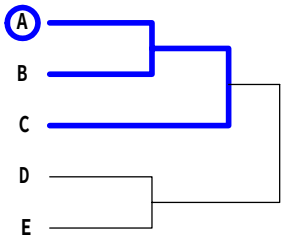

図 2　デンドログラムの例

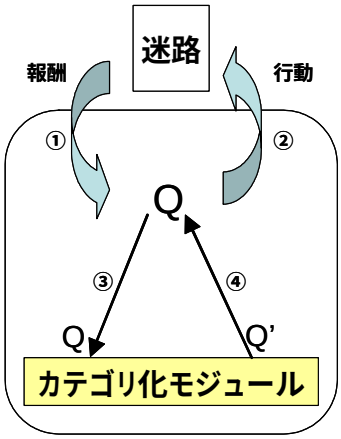

図 3　学習モジュールの周りとの関係

3. 提案手法

3.1 概要

提案する手法は，前章で述べたようにカテゴリ化と強化学習を組み合わせることによって行う．概要図を図 1 に示す．

本稿では階層型クラスタリング手法のウォード法と Q 学習を用いておこなう．また，学習対象は 3×3 の迷路を用いる．エージェントが迷路を解くことを迷路とエージェントとのインタラクションとする．

3.2 複数の学習器

本システムは図 1 のように対象に対してそれぞれ学習モジュールを持っている．

通常のひとつの学習器では，すべての知覚を状態として学習を行う．本稿で提案するシステムは，設計者が任意で，知覚のひとつを学習の状態から切り離し，その知覚ひとつひとつの状態に対して，学習器を選択できる．学習器の選択に用いられる知覚以外はそれぞれの学習器の状態として利用する．

本稿では，複数の迷路を効率よく解かせるため，迷路の違いが分かる知覚（迷路番号が分かる知覚）を学習から切り離し，迷路一つ一つに学習モジュールを与える．

3.3 カテゴリ化

それぞれの学習器は，対応する対象（迷路）とインタラクションをとり，学習結果を得る．この学習結果を図 1 の下部にあるカテゴリ化モジュールでカテゴリわけを行う．このカテゴリわけにはウォード法を用いる．ウォード法を用いると，図 2 のようなデンドログラムを生成することができる．このとき，下の階層とひとつ上の階層につながっているものを類似しているとみなし，その対象をカテゴリ内にあると考える．例えば，図 2 において太線（青）で示してあるように，A のカテゴリ内に含まれるものは B，C である．また，C のカテゴリ内に含まれるものは A，B，D，E である．

クラスタリング手法によって類似した対象の学習結果をみつける．それらをカテゴリ化モジュール内で平均し Q'を生成する．そして Q'を学習モジュールに送って利用する学習モジュール内の Q 値の更新をする．

Aのカテゴリ内にn個の要素があるときQ'_Aは以下の式で求められる．

$$Q'_A = (Q_B + Q_c + \cdots + Q_n)/n$$

3.4 学習の流れ

本システムは Q 値の更新に，通常の Q 学習の更新と，カテゴリ化モジュールで生成される Q'をもちいた更新の二通りある．

ある学習モジュールをひとつだけ取り出し，その周りとの関係を図 3 に示す．

図 3 の①から④に沿って順に説明をする．

①は迷路から報酬を受けることをあらわしている．②は Q 値を用いて行動を選択している．この①による Q 値の更新は次式になる．

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a')]$$

a は行動，s は状態，s'次の状態，α は学習率，γ は割引率である．本実験では α は 0.05　γ は 0.9 で行っている．

③はクラスタリングに Q 値をいれる．④はクラスタリング結果を融合させた Q'を用いて Q を更新している．更新式は次式になる．

$$Q = \beta Q + (1-\beta)Q'$$

β は定数で 2/3 で実験は行っている．

この Q と Q'の融合の割合は経験的に求めた．

またカテゴリ化モジュール④の更新の頻度は①②の行動報酬の更新 20 回，300 回に一回の割合で本実験は行っている．

4. 実験

4.1 実験概要

本実験は本稿で提案したシステムを組み込んだエージェントの動作確認および検証をするために行う．はじめに 30 種類の迷路を解いた後，その学習結果を用いて新しい迷路をどのように解いていくかを見る．

4.2 迷路の設定

3×3 のセルからなる迷路を 31 種類用意する．それぞれ迷路に番号が振ってある．左上に「S」右下に「G」と記述してあり，スタートとゴールを意味する．エージェントには明示しない情報として以下のものがある．スタートからゴールの経路は 6 種類存在する．それ以外の経路はランダムに生成してある．それぞれ 6 種類の迷路の代表例は図 4 の A から F である．31 種類の迷路は表 1 のように代表例に対応した迷路である．1～30 番の迷路が先に解いておく迷路で 31 番それを踏まえて新しく解く迷路である．

学習に用いる状態は迷路の位置とし，行動は上下左右の 4 つである．ひとつの迷路につき 32 個の行動が存在しそれに対する Q テーブルが存在する．行動を一回行うことを 1 ステップと数える．

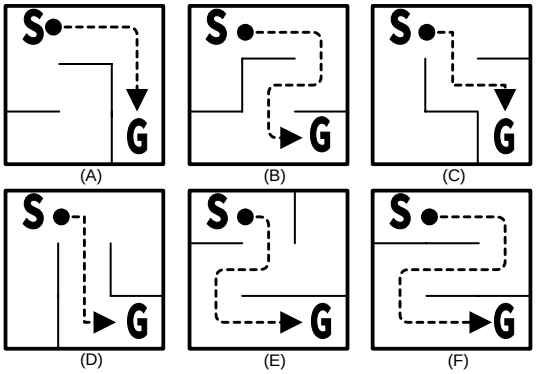

図 4 それぞれの迷路の種類の例

表 1 迷路の分類

迷路の種類	迷路番号
A	1〜9, 31
B	10〜12
C	13〜18
D	19〜27
E	28, 29
F	30

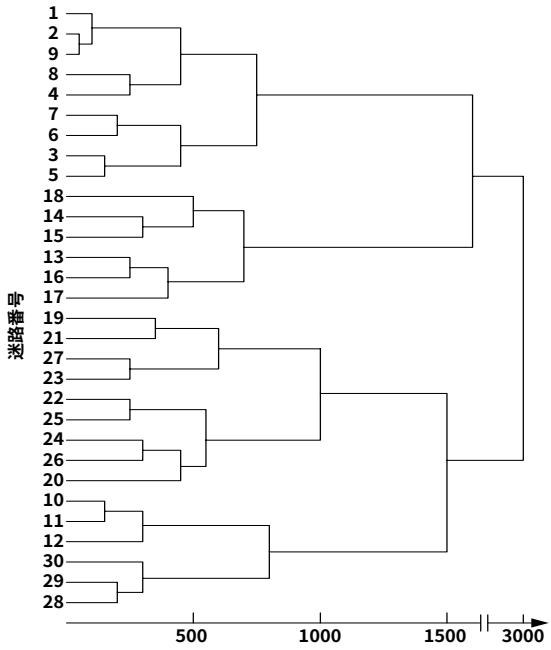

図 5 迷路のクラスタリング結果

4.3 実験手順

(1) クラスタリングを行い，デンドログラムを作成する

1 番から 30 番までの迷路を 2000 ステップずつ行動させ，十分に学習させる．その後 1 番から 30 番までに対応した Q 値をウォード法によりクラスタリングを行う．

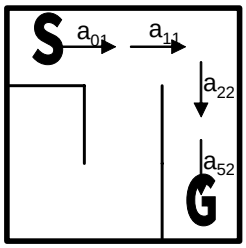

図 6 31 番の迷路の解の行動

(2) 新しい迷路(31 番)を解く(①から③それぞれ 2000 ステップの行動をみる)

① 初期値 2 の Q 値を用いて普通の Q 学習で迷路を解いた結果．

② クラスタリングをはじめる時期を，ステップ 150，ステップ 200，ステップ 300 とし，クラスタリングする周期を 20 ステップごとにして迷路を解いた結果．

③ クラスタリングを始める時期がステップ 0，ステップ 250 のものにおいて，周期を 20 ステップ，600 ステップに変化させて迷路を解いた結果．

4.4 実験結果

(1) クラスタリングからデンドログラムを作成する．

1 番から 30 番まで解いた Q 値によって作成したデンドログラムを図 5 に示す．

これを見てみるとスタートからゴールまでの経路で分類した A から F に対応するように分かれた．距離 1000 できってみると見てみると，上から 1 グループ目は A に対応している．2 グループ目は C に対応している．3 グループ目は D, E に対応している．4 グループ目は B, E, F に対応している．

(2) 新しい迷路を解く

新しい迷路(31 番)の解の行動$a_{01}a_{11}a_{22}a_{52}$を図 6 に示す．

それぞれの実験の結果は行動に対応した Q 値の変化をグラフに表す．

① 普通の Q 学習の結果を図 7 に示す．それぞれの Q 値は収束していることが分かる．

② クラスタリングの開始時期がステップ 0，100 のときはなかなか収束が始まらなかった．クラスタリング開始ステップが 200，300 のときの結果を図 8，9 に示す．それぞれ収束していることが分かる．

③ 250 ステップからはじめた 20 ステップ間隔と 600 ステップ間隔の結果を図 12，13 に示す．それぞれ収束していることが分かる．

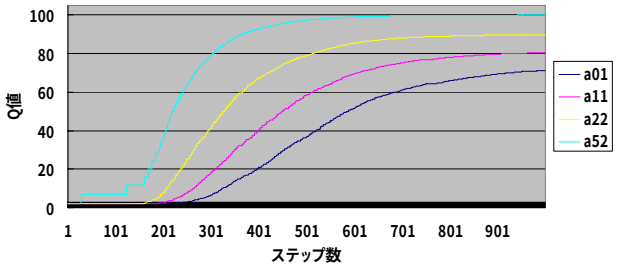

図 7　普通の Q 学習の結果

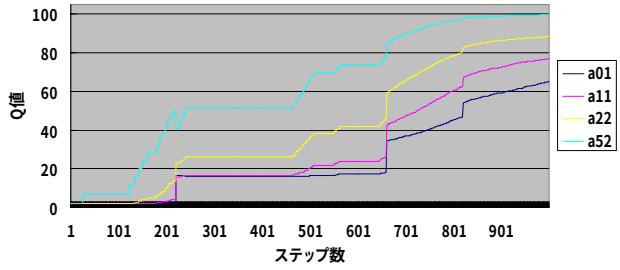

図 8　200 ステップからクラスタリングを開始した結果

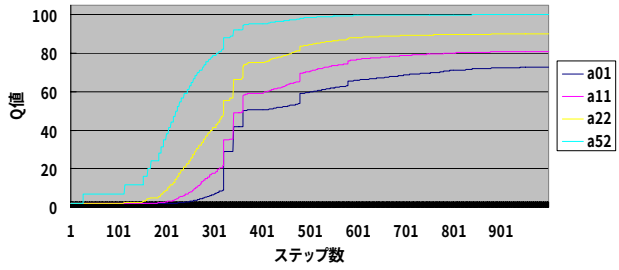

図 9　300 ステップからクラスタリングを開始した結果

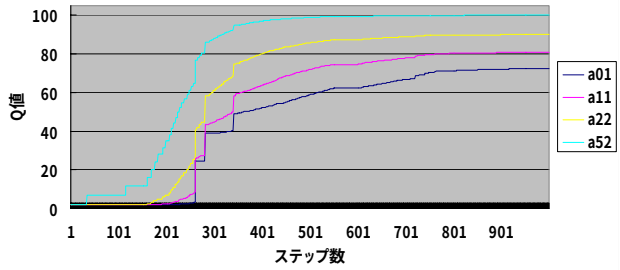

図 10　250 ステップから 20 ステップ間隔でクラスタリングした結果

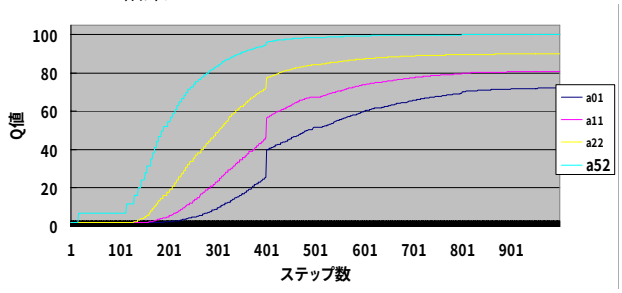

図 11　250 ステップから 600 ステップ間隔でクラスタリングした結果

4.5 実験考察

(1)クラスタリングでデンドログラムを作成した(図 5). 学習結果のデータでクラスタリングしたので, 近いものは経路が類似している. このデンドログラムからの類似した迷路を選んで, 仮想的に学習ができると考えられる.

(2)0, 100 ステップのような初期段階だと Q 値がほとんど立ち上がっていない. そのため, まったく性質の異なる迷路の Q 値を仮想経験として適用してしまっている. このような状態になると, 収束が困難になってしまうので収束が始まらないと考えられる. 200 ステップ(図 8)では収束しているものの, 全体として通常の Q 学習より遅れをとっている. これも, 0, 100 ステッのプ理由と同様と考えられる.

250 ステップ(図 10)や 300 ステップ(図 9)では普通のQ 学習と比べるとa_{52}, a_{22}はわずかに立ち上がりが早いが, a_{01}, a_{11}は明らかに早い立ち上がりになっている. これは類似した迷路のQ 値が有効に使えた結果であると考えられる.

図 12, 図 13 を見てみるとステップ間隔が小さいとゴールから遠い経路の Q 値の学習が早くなる. 大きいときはほとんど通常の Q 学習を行い 2 回だけ効率よく Q 値を更新している.

全体を通して, クラスタリングを行う時期, クラスタリングを行う回数が学習の早さに関わってくることが分かった. ある程度, 迷路の特徴を学習してから出ないと, 間違った仮想学習をしてしまうことが分かった.

5. まとめ

学習アルゴリズムにクラスタリング手法を付加することで, 仮想的に学習するシステムを提案した. そしてその動作について確認をした. 現段階では, パラメータの設定を人間の手でしっかりしないと学習の促進にはつながらない. しかし, このパラメータも自ら学習するようになれば, 過去の経験から突然賢くなったり, いつまでたっても成長しなかったり, といった予想に反する行動は人とのインタラクションにおいて, 重要な振る舞いにつながっていくのではないかと考えている.

参考文献

[山田 02] 山田誠二, 角所考: 適応としての HAI, 人工知能学会誌 17 巻 6 号, pp658-664(2002)

[片上 05] 片上大輔, 大村英史, 安村禎明, 新田克己: 社会的インタラクションに基づくマルチユーザ学習エージェント(MULA), 日本知能情報ファジィ学会誌, (2005.5)

[片桐 03] 片桐恭弘: ロボットの社会的知能, 情報処理, Vol.44, No.12, pp.1233-1238(2003)

[中島 04] 中島宏, 森嶋泰則, 山田亮太, Scott Brave, Heidy Maldonado, Clifford Nass, 川路茂保: 人間－機械協調システムにおける社会的知性, 人工知能学会論文誌, 19 巻 3 号, pp184-196(2004)

[山岸 03] 山岸栄輝, 塩瀬隆之, 川上浩司, 片井修: 学習空間のモジュール間インタラクション, 人工知能学会 第 4 回若手の集い MYCOM(2003)

[鮫島 02] 鮫島 和行,片桐 憲一,銅谷 賢治,川人 光男:電子情報通信学会 論文誌, J85-D-II, 90-100 (2002)

[Reeves 96]Byron Reeves, Clifford Nass: The Media Equation, Cambridge University Press (1996)