コミュニケーション相手の取捨選択による個体知能の効率的発達

認知ロボティクス研究室　　木島 康隆

2010 年 2 月 9 日

1　研究背景・目的

近年，ロボットは高機能・複雑化してきた．それに伴ってロボットの適用範囲も家庭用やエンターテインメント用途など多方面に広がっている．このような中でロボットに要求されていることとして，人間の生活空間のような多様で動的な環境の中でも使用者の用途に応じた適切な行動をとることが挙げられる．これに対して，機械学習手法を用いロボットに環境に適した行動を学習させる研究が行なわれている [1][2]．ロボットの行動学習に関する研究は数多く存在するが，本研究では郡ロボットを用いた個体知能の発達に関する研究に注目する．群ロボットを用いた研究では，個体間でのコミュニケーションによる協調学習や，競合学習などによって環境に対して適切な行動を学習させる研究が行なわれている [3][4]．

コミュニケーションは主に通信により行なわれ，群中の個体はコミュニケーションにより入手した他者情報と自己の試行錯誤による経験を基にして知能を発達させる．しかし，どのような個体とコミュニケーションするかといった設定は人間が行なってきた．つまり，人間の主観で有益であると考えられる個体とコミュニケーションするように設定していた．そのため人間の設定が不適切であった場合，有害な情報がやりとりされ，学習効率が下がってしまうことが想定される．この問題に対して，ロボットが自身にとって有益な情報をもたらす個体を選定し，コミュニケーションを行なうことで効率的な学習を行なうことを考える．よって，本研究の目的は効率的な個体知能の発達のために，コミュニケーション相手の取捨選択を行なうシステムの構築である．

2　アプローチ

ロボットがコミュニケーションに有用な他者の選定方法についての概念を述べる．今回は，実際に他者とのコミュニケーションを通して有用なコミュニケーション相手を学習させることで他者の取捨選択を実現することを考えた．図 1 に概念図を示す．図 1 より他者から提供された情報をもとに行動を行い，行動の結果に基づいて情報提供した他者を評価する．行動の結果が，自身にとって有益なものであれば，情報提供した他者は自身にとっても有益な情報をもたらす個体と考え，評価を高める．さまざまな他者とコミュニケーションを行なうことで，それぞれの他者に対する評価が決定されていく．自身は評価の高い個体とコミュニケーションをとることで有益な情報のみを受け取りより効率よく学習することができる．このような他者とのコミュニケーションを通して自身に有用な情報を持つ他者を学習し，積極的にコミュニケーションを行なうシステムを提案する．

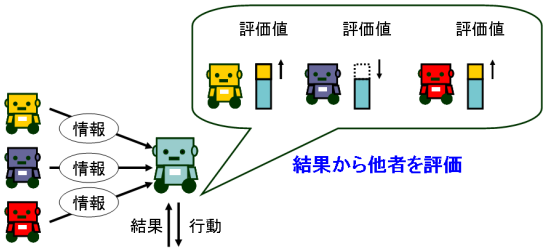

図 1　アプローチ

3　提案システムの概念

本研究では強化学習 [5] を用いてシステムを構築する．手法は 2 つの学習機構に分けられる（図 2）．1 つは，今回考える自身にとって有用な他者の選定を学習する他者選択学習部，もう 1 つは直面する状態に対して適切な行動を学習する行動学習部である．他者選択学習部は，他者学習と他者選択の 2 つの機能から構成されている．他者選択はコミュニケーションを行なう他者を決定し，選択した他者に対して情報の要求を行なう．このとき他者を決定する判断材料となるのが他者知識である．他者知識は，ある他者が提供する情報がどれほど有用なのかという情報である．他者知識は他者学習部分によって更新される．他者学習では行動学習部での選択した行動に対して受け取る報酬を基に他者の情報が正しいかどうかを評価し，知識とする．行動学習部は，行動学習と行動選択の 2 つの部分から構成されている．行動選択は，直面する環境状態に対する行動を蓄えられている行動知識と他者からの情報を基に選択する．行動知識は，ある環境状態である行動をとることがどの程度よいものかという情報である．行動学習は行動した結果，環境から受け取る報酬から行なった行動の評価を行い行動知識とする．行動選択と行動学習を繰り返

すことで行動知識を蓄えていく．本論文では，行動の度に報酬が与えられる即時報酬環境に対応する手法とエージェントが目的を達成したときに報酬が与えられる遅延報酬環境に対応する手法の 2 種類の手法を提案する．しかし，本稿では紙面の都合上，遅延報酬環境についてのみ解説する．

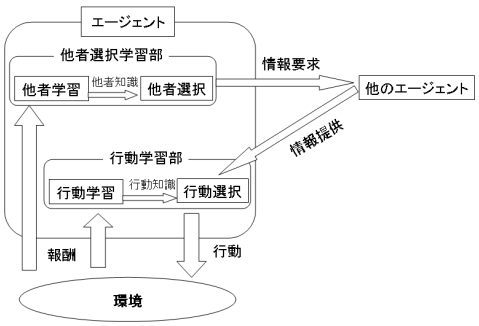

図 2　システムの概念図

遅延報酬環境に対応する手法について概要を説明する．遅延報酬環境環境ではエージェントが目的を達成したときに報酬が与えられる．そこで，より目的達成に近い時点でコミュニケーションを行い，情報を採用した他者ほど高い評価を与えるようにする．したがって，目的達成までの各時点で他者の情報の採用記録をとり，目標達成時に他者の採用記録と報酬を基に他者の評価を行う．このとき，過去に採用したものほど割り引いて評価を行う．

4　実験

本論文では，迷路問題を対象タスクとして実験を行った．本実験で用いた迷路環境は，ゴールまでのルートが複数あるような迷路である．また，ゴールも 1 箇所ではなく複数箇所に設定する．エージェントは複数設定したゴールからランダムに 1 箇所割り当てられる．また，エージェントには寿命が設定したあり，寿命に達した個体は迷路環境から取り除かれる．そして，取り除いたエージェントの代わりに新しいエージェントを迷路環境に投入する．これにより，迷路環境内には学習の進んだ個体とそうでない個体が共存する．以上のような環境のもとで実験を行った．

実験結果を図 3, 図 4 に示す．図 3 は群全体での生涯獲得報酬量の総和である．この結果は，群単位でみた場合の手法のパフォーマンスを示している．図 4 は 1 個体あたりの生涯獲得報酬量の平均を示している．いずれの結果も提案手法が高い獲得報酬を示している．

5　結論

本研究では，コミュニケーション個体の取捨選択により個体知能をより効率的な発達を実現するシステムを提案した．

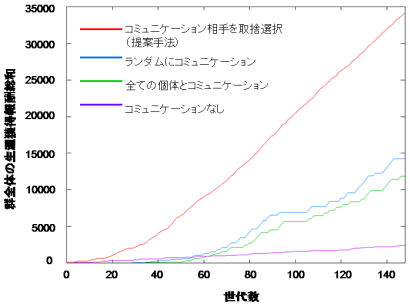

図 3　群全体の獲得報酬総和

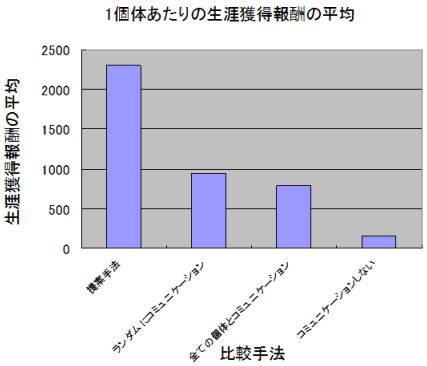

図 4　1 個体あたりの生涯獲得報酬平均

検証実験として迷路問題に適用し，獲得報酬量の視点から手法の有効性を確認した．

参考文献

[1] Jun Tani,“ Model-based learning for mobile robot navigation from the dynamic system perspective ”, IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics, Vol.26, No.3, pp.421-436, 1996.

[2] Y. Takahashi, M. Asada, “ Multilayered learning systems for vision-based behavior acquisition of real mobile robot ”, Proceedings of SICE Annual Conference 2003 in Fukui, pp.2937-2942, 2003.

[3] Ian D. Kelly, David A. Keating, ”Increased Learning Rates Through the Sharing of Experiences”, Proceedings of the Seventh IEEE International Conference on Fuzzy Systems, 1998

[4] Yasutaka Kishima, Kentarou Kurashige, ”Growth of individual intelligence using communication”, Proceedings of SCIS & ISIS 2008, pp.287-292, 2008.

[5] Richard S. Sutton, Andrew G. Barto, ”Reinforcement Learning”, The MIT Press, 1998.