3B2-04

マルチユーザ学習エージェントによる社会的インタラクションの効果

Effect of Social Interaction using Multi User Learning Agent

片上 大輔 Daisuke KATAGAMI
新田 克己 Katsumi NITTA

東京工業大学大学院 総合理工学研究科
Department of Computational Intelligence and Systems Science
Tokyo Institute of Tokyo Technology

In this paper, we propose Multi User Learning Agent (MULA) based on *Reinforcement Learning* to interact with various types of people effectively towards realization of Social Interaction. MULA is equipped with two functions for social learning. One is direct learning function by individualization of user parameters, and another is indirect learning function by past experience and the similarity between each users. We verify the effect of social interaction of social learning system Multi User Learning Agent (MULA) influenced according to the similarity between users using the computer simulation.

1. はじめに

近年ロボットやエージェントの社会性の獲得に関する研究が注目されている。この分野において、[Fong03]は、このテーマに関する貴重なサーベイである。ここにも、さまざまな社会性を目指したロボット研究が紹介されているが、現状では基本的に1対1における社会性を対象とした研究が多い。方法論としても人間のような社会的能力を設計者が埋め込み的に設計するアプローチが多く、そこでは広く人間とインタラクトするような広義の意味での社会性の実現を目的としているように思われる。

一方、人工物(エージェントやロボット)における知性の設計は、従来 AI の分野で行われてきた記号処理に基づく人工知能では限界があると言われている。近年のさまざまな研究により、身体性や、環境との相互作用に基づく知性の構築の重要性が認識されてきた。「認知発達ロボティクス」に代表されるように、構成論的アプローチによる知性の獲得が重要であり、人工物の真の人間らしさを目指すには、これらのアプローチのようにボトムアップ的に迫る必要がある.

前述のように、自律的な人工物が真の社会的知性を獲得するには、機械学習(machine learning)に代表されるような学習機構に、周りの主体との相互作用から自己組織化的に自己の行動を学習するようなボトムアップな学習機能(これを本稿では社会的学習とよぶ)を組み入れる必要があると考える。このようなボトムアップに創出する狭義の意味での社会性獲得にこそ、人工物における普遍的な人間らしさやそこにおける個性が生まれるものと筆者らは考えている。

本稿では、直接的経験とは別にユーザ間の類似性に応じて影響を受けるといった代理経験的な学習を行うエージェントである、マルチユーザ学習エージェント(MULA)において計算機上におけるマルチタスクシミュレーションによりその社会的インタラクションの効果を検証する。

2. 社会的環境と学習機能

学習という視点で、社会的環境を想定すると、以下のような幾つかの問題点が浮き上がってくる。

1. 複数の報酬源の存在: 異なった多くのユーザから報酬が与えられるため、ユーザによって矛盾していたりすることもある。これをどう統合し、どう学習させるのかが難しい問題となる。
2. ユーザからの報酬の不一致性: ユーザには様々な特性(嗜好、意図など)があり、一般にそれは一致していない。事前知識によりわかることも難しく、相互作用を通して個別に獲得する必要がある。
3. 各ユーザの報酬の変動性: 各ユーザからの報酬もインタラクションを長く続けているとユーザの特性(熱心さ、飽き易さ、嗜好、感情の変化など)により同じ状況においても報酬が変動してしまう。これが学習の収束を難しいものにしている。
4. 訓練の希薄性: 現実には、各ユーザと十分な訓練を行うことは困難であり、十分な訓練を想定することは現実的に有用なシステムとはいえない。このスパーシティ(sparsity)の問題を克服する必要がある。

これに対して、我々はボトムアップ的に学習する社会的エージェントに必要な機能を以下のように考える。

A) 対応しているユーザの同定: 複数ユーザとの知的なインタラクションのためには、現在誰と対応しているのかを同定することが重要である。(1. に対応)
B) 適切な状態空間の構築: ユーザ間の類似性に応じて影響を受ける社会的学習においては、同定したユーザの特性を適切に構築しなくてはならない。この精度が学習の精度に直接影響する。(2. と 3. に対応)
C) ユーザ間の類似性の計算: 構築した各ユーザの状態空間の比較により、類似性を計算する。これも、状態空間の何に注目して類似性を測るのか、または類似度計算の精度が学習の精度に影響する。(2. と 3. に対応)
D) 類似性に応じた各ユーザへの対応の学習: 類似性の大きさに応じて代理経験的に学習が行われる。つまり実際の経験により学習が行われる時に、現状と似たような状況の時にも対応できるように学習が行われる。これにより一つの実経験を行う際に頭の中で仮想的に経験をつむことができ、希薄性の問題への解決のアプローチとなる。(2.と 3.と 4. に対応)

従来の多くの社会的研究では、上記の問題に対して、人間の社会的能力を個々に直接設計するアプローチであった。我々はこれらの4つの機能をうまく用いることによりボトムアップに社会的学習を行うことが可能になると考える。本稿ではまずこの中でも最も重要な D)に関して、実装と検証を行う。

連絡先:片上大輔, 東京工業大学大学院総合理工学研究科, 〒226-8502 横浜市緑区長津田町 4259, 045(924)5218, katagami@ntt.dis.titech.ac.jp

3. マルチユーザ学習エージェント:MULA

3.1 概要

前述の社会的学習機能を備えたエージェントを本研究では，マルチユーザ学習エージェント（Multi User Learning Agent: MULA)と呼び，モデルベース強化学習(Dyna-Q アルゴリズム[Sutton 90])を基本として，以下に説明する2つの学習方法によりMULA を実現する.

(1) パラメータの個別化による直接的学習

マルチタスク学習では，複数の環境に対応するために環境毎に知識やルールセットを構築しそれを再利用するアプローチが多く用いられる[Parr 98]. しかし，それぞれの環境に対しそれらを用意するのは効率が悪く，メモリの問題やそれらの知識の重みづけの問題などが発生する. そこで，知識は一貫して同じものを用い，知識の用い方(パラメータ)に関してだけユーザ毎に個別に用意するアプローチを採用する.

(2) 過去経験による間接的学習

複数のユーザに効率的に対応するためには，複数のユーザとの過去経験をうまく利用した社会的な対応が重要である. 対象ユーザとのインタラクションは過去の類似ユーザとの経験が役に立つ. 特に対象ユーザとの経験が浅い段階では，インタラクションの指針になりうる. そこで，ユーザ間の類似性を用いて，対象のユーザと類似の過去のユーザとの経験を利用する.

3.2 MULA の学習手続き

MULA アルゴリズムを図1に示す. MULA ではまずユーザを特定する. そのユーザ A とのインタラクション(ある状態 S_A に対するユーザへの行為出力 a_A に対して報酬 r_A を受け取る)から simple Q-learning を用いて，ユーザに個別の政策/価値関数を直接的に更新する. 次にユーザの情報を用いてユーザ A と他ユーザ X の User Profile の類似度 S_{AX} を計算する. この類似度とユーザ A とのインタラクションの情報 S_A, a_A, r_A を利用して，間接的に各ユーザ全員の User Profile を更新する. この 2 つの経験により前節の直接的学習，間接的学習を実現する. 本研究では，q 値の更新に社会的要素を導入している. この q 値の更新は図中の最後の式に対応している。ここで，$q(s,a,j)$ はユーザ j に対する状態 s、行為 a における q 値，α は係数，r_A はユーザ i からの即時報酬，$r_{\max}$ は最大報酬である。ここでは，ユーザ i との経験により間接的にユーザ j の予測値を更新している.

4. 計算機実験による手法効果の検証

4.1 実験目的

3 章の間接的学習を利用した社会的強化学習法を実装して、スパーシティ問題に対する性能検証を行う。

4.2 実験設定・計画

実験のタスクとして図 2 に示すようなグリッドワールド上における迷路問題を用いた。これは、Sutton が Dyna システムの検証のために作成したものであり、ベンチマーク的問題となっている。始点(S)から終点(G)までの経路を学習する問題である。6x9 の各セルに7つの障害物があり、46+終点(G)の 47 状態が存在する。エージェントは常にその中のどれか1状態(セル)にあるものとする。エージェントは東西南北の 4 方向への移動が可能であり、決定的に状態遷移が行われる。エージェントは障害物や

```
Initialize q(s,a,i) for all s,a,i
Do forever
   User identification i ∈ I
   Simple Q-learning
   User Profile R_i(s,a,i) ← r_i
   For j=0 to j=n do
      Similarity S_ij ← User Profile R_i, R_j
   q(s,a,j) ← q(s,a,j) +
   αS_ij(| r_i + γ max_a' q(s',a',i) − q(s,a,i) |)
```

図 1　MULA algorithm

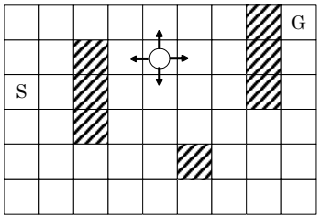

図 2　迷路問題 (Sutton)

枠外へ遷移することはできない。終点状態(G)に達すると報酬値 100 が与えられて始点(S)へと戻される。あらかじめ終点(G)の位置を知らないエージェントは、終点で得られる報酬 と試行錯誤により自律的に最短経路を獲得することを目的とする。

ここでは、ユーザをこの迷路(タスク)に見立てて、エージェントに対して複数の迷路問題が順番に与えられるという設定を考える。文献[田中 03]に倣い、上述の Sutton 迷路では決定的であった状態遷移を確率的なものに変更し、その確率値に変動を持たせることにより各タスク同士に違いを作成する。本実験ではユーザの報酬の不一致性への対応として、簡易的に確率値をある正規分布から作成して、その正規分布の集合を形成する。エージェントに新たにタスクを与える際には、その集合を用いて全ての状態遷移確率値を独立に作成し、一つのタスクを作成する。

5. おわりに

社会的環境に潜在する社会性の構築における問題点と、そこへのアプローチ、特に過去経験による代理経験的な社会的学習方法について述べた。今後、社会的環境をマルチタスク環境にみたてて計算機シミュレーションにより検証する予定である。

参考文献

[Fong 03] T. Fong, I. Nourbakhsh and K. Dautenhaln: A survey of socially interactive robots, Robotics and Autonomous Systems vol. 42, pp.143-166, 2003.

[Parr 98] R. Parr and S. Russel: Reinforcement Learning with Hierarchies of Machines, Advances in Neural Information Processing Systems 10, pp.1043-1049, 1998.

[Sutton 90] R. S. Sutton: Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming, Proc. of the 7th International Conference on Machine Learning, pp.216-224, 1990.

[田中 03] 田中文英, 山村雅行: MDP 集団の上におけるマルチタスク強化学習，電気学会 電子・情報・システム部門誌, vol. 123, no. 5 , 2003.