4B1-R-2-2

行動モデル学習における動的環境の影響

Behavior Model Learning under Dynamic Environment

市瀬 龍太郎[*1] 森山 甲一[*2] 沼尾 正行[*2]
Ryutaro Ichise Koichi Moriyama Masayuki Numao

[*1]国立情報学研究所 大阪大学 産業科学研究所[*2]
National Institute of Informatics ISIR, Osaka University

We propose a machine learning method for generating behavior model. The method utilizes multiagent simulation approach. We conducted experiments with real simulation environment. The experimental results show that the performance of the proposed method is better than previous method.

1. はじめに

人間は，さまざまな環境をセンシングし，行動の決定を行う．それと同じようなことを実現するために，人間の行動履歴から行動を模倣するプログラムを作成する，行動クローニング [Sammut 96] の研究が行われてきた．しかし，実環境において，人間の行動は，すべての選択肢をカバーしているわけではない．本論文では，シミュレーション環境を用いることで，実際の人間の行動には表われないような行動についても，妥当性を調べながら，行動モデルを自動で学習する手法について述べる．そのような手法として，従来までは，シミュレーションの結果に応じて，行動を随時変更し，学習する手法が用いられていた [市瀬 11]．しかし，シミュレーション環境が動的な場合においては，シミュレーション結果と同等の結果が常に得られるとは限らないため，シミュレーションの結果を利用する際には，環境の影響を考慮する必要がある．本論文では，この手法を様々な環境で試験することによって，手法の環境適性を明らかにする．具体的には，シミュレーション手法を変更することにより，より正確な行動を学習する手法について述べる．

2. 対象とするシミュレーション環境

本研究では，対象とするシミュレーション環境として，Happy Academic Life 2006(HAL2006) [山川 06] というゲーム型キャリアデザイン学習教材を用いた．HAL2006 は，人工知能学会 20 周年記念事業として開発された教育用ボードゲームで，プレイを通して，研究者のキャリアデザインを学習できるようになっている．プレイヤは自分のコマを進めながら，さまざまなイベントを疑似体験し，研究業績を積み上げて最終的なゴールを目指す．ゴールには，教育者型，悠々自適型，学内政治型，学術社会型，業績量産型，組織研究型，業績卓越型の 7 つがある．学習者は，プレイ途中で，体験するイベントにどのような判断をするかによって，自分の置かれる状況が変化する．そのため，プレイヤはさまざまな場面において，自分のゴールを達成するための適切な判断をしなければ，ゴールになかなか到達できないことになる．

HAL2006 は, 当初，紙を使ったボードゲームとして開発された．それを研究プラットフォームとして再構築し，電子化を行ったものが D-HAL2006(図 1) [市瀬 08] である．D-HAL2006 では，複数の人間の学習者が計算機を使ったプレイで，学習できるのみならず，人間の思考と同様な行動ルールを記述することで，人間の代わりに，エージェントがプレイすることもできるようになっている．本研究では，このシミュレーション環境を用いて，なるべく早くゴールすることができるような行動モデルを学習することとする．

図 1: D-HAL2006

3. シミュレーションによる戦略学習

これまでに，遺伝的アルゴリズム (GA) を用いた進化計算によって，上記のゲームにおける戦略を獲得する手法が開発されてきた [市瀬 11]．その手法では，戦略ルールはそれぞれ遺伝子の列で表現され，各個体はその組合せからなっている．個体間の交叉と突然変異により複数の新個体を生成してシミュレーションを行い，その結果に基づいて個体を選択することを繰り返すことで，より早くゴールできる個体（ルールの組み合わせ）を発見する．しかし，本研究で想定しているシミュレーション環境は，動的な環境となるため，同じシミュレーション結果がいつも得られるとは限らない．例えば，サイコロを振るという一点をとっても，常にシミュレーション時と同じ目がでる訳ではないため，ある戦略によって，シミュレーションがうまくいったとしても，実際に使う機会にうまくいくとは限らない．つまり，環境の変化に弱いという問題点がある．そこで，本研究では，シミュレーション環境の評価回数を増やして安定化させると同時に，多くのシミュレーション環境を経た個体を優先的に利用することで，より正確な戦略獲得を試みる．

連絡先: 市瀬 龍太郎，国立情報学研究所情報学プリンシプル研究系，〒 101-8430 東京都千代田区一ツ橋 2-1-2，Tel:03-4212-2000，E-mail:ichise@nii.ac.jp

4. 実験

まず，5 回のゲームでゴールするまでのターン数の合計を適合度とした GA で，100 個体 100 世代による実験を行った．ここでは，ゴールの種類によって戦略が異なるため，ゴールの種類毎に，行動モデルを作成することとした．得られた最良の行動モデルを用いて 100 回ゲームを行ったところ，悠々自適型，業績卓越型の 2 種のゴールに対しては，平均で 50〜60 ターンでゴールすることができた．これは 1 回のゲームでゴールするまでのターン数を適合度とする従来手法 [市瀬 11] と同様の結果である．しかし，ゴールが業績量産型の場合は，従来手法の平均が 85 ターンであったのに比べ，本研究の手法では 56 ターンと大幅に改善している．また，組織研究型については，従来手法では 100 回のうち，24 回でゴールまで到達できなかったが，本研究の手法では全ててゴールまで到達できており，そのターン数も 91 ターンから 77 ターンへと改善している．その他の教育者型，学内政治型，学術社会型については，従来手法とほぼ同様の結果（70〜90 ターン）であった．

次に，平均で 60 ターンを超えている教育者型，学内政治型，学術社会型，組織研究型の 4 つのゴールについて，300 世代まで実験を行ったところ，平均ターン数は教育者型で 91 から 71 と減少するが，その他の 3 ゴールでは結果は大きく変わらなかった．ところが，縦軸に頻度を取り，横軸をゴールまでのターン数を 10 ターン毎のまとめると，興味深いことが分かった．それを示したものが，図 2 である．ここでは，紙面の都合上，教育者型と学内政治型だけ掲載している．色の違いは，世代数を示している．この図は，正規分布型よりもロングテール型に近い形状となっている．つまり，大部分のゲームは少ないターン数でゴールできるが，一部で非常に大きなターン数を要していることを示している．そこで中央値をみると，4 つのゴールのいずれの場合でも，中央値は平均ターン数よりも小さな値となった．特に学内政治型では，平均 77 に対して中央値 66，学術社会型では，平均 89 に対して中央値 83，と差が開いている．また，学術社会型以外では，中央値は 73 ターン以下となっていた．この結果より，多くのゲームにおいて，良い結果をもたらす行動モデルが学習できたが，その行動モデルには，ある条件がそろった特殊な環境において大きく失敗するという弱点があると言える．しかし，図 2 より，世代数が増加するにつれて，分布が左に移動することが読み取れるため，さらに世代数を増加させることで，この弱点を解決することができると考えられる．

5. おわりに

本研究では，シミュレーションを取り入れた行動モデル学習手法について述べた．シミュレーションを用いる場合には，人間の行動履歴から行動モデルを学習する手法に比べて，人間が遭遇しなかった状況に対しても学習を行うことができるという利点がある．しかし，シミュレーションが動的な環境の場合には，シミュレーション結果自身がバイアスを受ける．それを軽減させるために，本研究では複数のシミュレーション結果を統合して利用することを提案した．その結果，行動モデルが改善されたり，多くの場合によい結果をもたらす行動モデルが学習可能となった．

この研究では，シミュレーション環境に対する知識は，全く用いていない．しかし，現実問題では，シミュレーション環境に対する何らかの知識を人間があらかじめ持っていることが多い．今後は，人間の環境に対する知識を取り入れることで，多様な環境要素と効率的な行動との因果関係の同定を行い，本研究で開発した手法と統合することで，さらに正確な行動モデルを学習できるようにしたい．

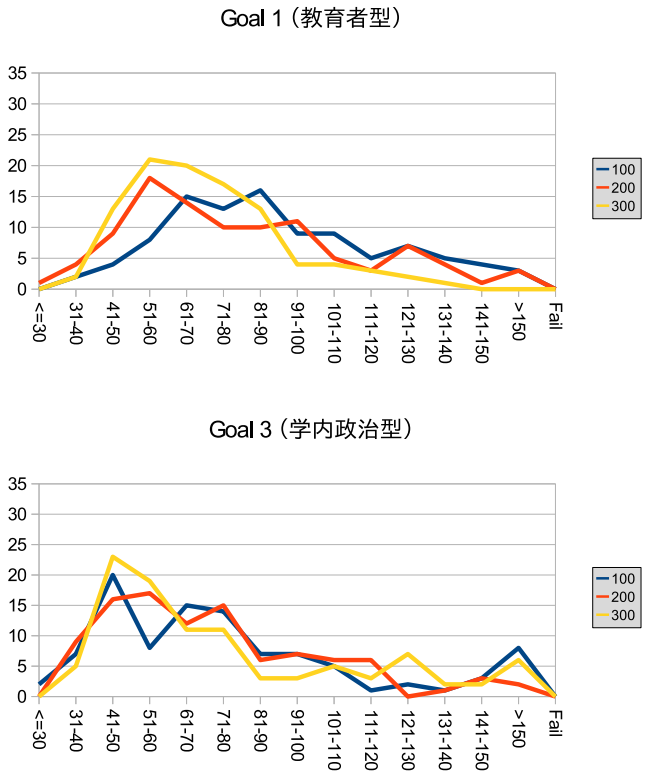

図 2: 教育者型と学内政治型のゴールまでのターン数の分布．横軸：ターン数，縦軸：頻度．

謝辞

本研究の一部は，物質・デバイス領域共同研究拠点における共同研究の支援により行われたものである．

参考文献

[Sammut 96] Sammut, C.: Automatic construction of reactive control systems using symbolic machine learning, *Knowledge Engineering Review*, Vol. 11, pp. 27–42 (1996)

[市瀬 11] 市瀬 龍太郎, 森山 甲一, 沼尾 正行：シミュレーション環境を用いた適切な行動モデルの学習, 第 25 回人工知能学会全国大会, 1G1-5 (2011)

[山川 06] 山川 宏, 市瀬 龍太郎, 太田 正幸, 加藤 義清, 庄司 裕子, 松尾 豊：Happy Academic Life 2006:研究者の人生ゲーム – ゲーム型キャリアデザイン学習教材の開発 –, 人工知能学会誌, Vol. 21, No. 3, pp. 360–370 (2006)

[市瀬 08] 市瀬 龍太郎, 庄司 裕子, 山川 宏, 三浦 麻子：学習者モデリング技術を用いたゲーム型教育システムのための研究プラットフォームの構築, 第 22 回人工知能学会全国大会, 2P2-12 (2008)