

BDI エージェントによる連続的な仮想世界におけるシミュレーションの実装

Implementing simulation environment for BDI agents in continuous virtual world

亀村 美佳^{*1} 新出 尚之^{*2} 柚木 静香^{*3} 宮田 怜奈^{*3} 高田 司郎^{*4}
 Mika Kamemura Naoyuki Nide Shizuka Yunoki Reina Miyata Shiro Takata

*1奈良女子大学 大学院 人間文化研究科

Graduate School of Humanities and Sciences, Nara Women's University

*2奈良女子大学 研究院 生活環境科学系

Faculty, Division of Human Life and Environmental Sciences, Nara Women's University

*3奈良女子大学 理学部 情報科学科

Faculty of Science, Information Science, Nara Women's University

*4近畿大学理工学部

Faculty of Science and Engineering, Kinki University

In recent years, research for the realization of robots which act autonomously in the real world is underway. In the continuous real world, we are faced with various problems different from ones in the virtual world. To deal with those problems, simulation of actions in the continuous world is desired. In this paper, we describe an implementation of simulation in continuous virtual world using canoe racing as a testbed. In designing the simulator, we separate the problem dependent part and the others, since we aimed a generic simulator that can handle the action by robot or agents in continuous world without depending on the specific task of canoe racing. Besides, we discuss introducing multiple method of reinforcement learning, since, to realize autonomous behavior, it is required to use actions acquired in learning for planning and the execution of the plan corresponding to individual tasks.

1. はじめに

我々は、実世界において自律的に振る舞うロボットに関する研究課題に取り組んでいる。連続な実世界では、仮想世界とは異なる問題に直面するため、それらに対処できる能力が求められる。これに際し様々な実験が必要となるが、実世界で直ちに実行に移すのは難しい。そこで、連続的な仮想世界におけるシミュレーション環境の実装を目指した。仮想世界におけるBDIモデルの有用性が[高田 12]で示されていることから、我々はBDIエージェントを利用したシミュレータを提案する。

これまでに我々は、カヌーレーシングをテストベッドとしたシミュレーションの実装に取り組んできた。

実世界は常に動的に変化しており、環境の変化を受ける。例えば、カヌーレーシングでは動的に変化する水流の影響を考慮する必要がある。また、離散的なグリッドワールドを用いた仮想世界とは異なり、エージェントの行動は正確に表現できず、誤差を伴う。シミュレータの設計に当たってはそれらを考慮し、エージェントによる実世界での自律的な問題解決をシミュレートできることを目指した。

実世界の完全なシミュレーションは難しいが、シミュレーション環境を次第に実世界に近づけていくことによって、有用な知見が得られると考えている。そこで我々は過去に、まず、任意の基本行為を繰り返して川を下るのみのシミュレーション作成に着手し、次に、エージェントに知覚と強化学習の機能を追加した。これにより、エージェントの最適な行動決定が可能となり、効率的な問題解決に至った。しかしながら、これはテストベッドであるカヌーレーシングのみに対応したものであり、更には、川の形も固定化されており、汎用性が乏しいものであった。

本研究ではこれを改善することで、カヌーレーシングに依存したのではなく、環境が異なる様々な問題設定に対応できるシミュレーションを目指した。

また、実世界でエージェントが目標達成のために行動できるためには、そのための技能を学習で獲得し、それを基本行為としてプランを選択して実行できることが求められる。我々のシミュレータは、BDIエージェントを用いることにより、学習で獲得した行為を、プランを構成する基本行為として利用し、熟考でのプランの選択による行動決定と併用できるようになっている。ここで、実世界での様々な問題を解決するには、問題に応じた柔軟な学習方式を取ることが必要である。そこで我々は、異なる学習方式を併用可能とすることで、エージェントによる行動決定までの過程を多様化した。

エージェントが行為の結果を学習し、問題解決のプランを作成及び実行するにあたり、より効率的なプランを選択するためには、複数のプランを考慮する必要がある。学習方式を多様化することは、プラン選択の幅を広げることができ、問題解決における最適な手段の導出を目的としている。

2. カヌーレーシング問題

本節ではシミュレーションにおけるカヌーレーシングの概要と問題設定を示す。

2.1 基本行為

カヌーカヤックは両側にブレード^{*1}があるパドルを使った、「フォワード・ストローク(前進)」、「リバース・ストローク(後進)」、「ストップピング(停止)スweep・ストローク(回転)」、「ドロー・ストローク(手前横方向移動)」、「スカーリング・ストローク(横方向移動)」と呼ばれる一般的に使用されているストロークがある[高田 13, 辰野 05]。本研究ではこ

連絡先: 亀村美佳, 奈良女子大学大学院情報科学専攻, 奈良市北魚屋西町, 0742(20)3555, kamemura-mika1015@ics.nara-wu.ac.jp

*1 水をキャッチする部分のことを指す。

れを踏まえて、基本行為を左右のストロークによる 5 つに定義した。すなわち、「前進」、「後進」、「左進」、「右進」、「何もしない」である。左進及び右進は、片側だけのストロークによる動作を想定したものである。

2.2 課題設定

エージェントは、連続空間内で基本行為を繰り返すことで目標の達成を目指す。本研究におけるカヌーレーシングの課題設定は、連続空間の中にある川の上端からスタートして下端に到達し、川を下り終えることである。またその上で、目標達成までに川から外れず、より早く下り終えることを目標とする。エージェントの行為の結果は、漕ぎ方と川の流れに影響を受け、また、同じ流れで同じ漕ぎ方をしたつもりでも、移動先には誤差を生じる。

3. シミュレーションの設計

1 節で述べたことを考慮し、シミュレータの設計においては、汎用性の確保と学習方式の多様化について考慮した。本節ではそれらについて述べる。また、これに際し本研究におけるシミュレーションの特徴を述べる。

3.1 シミュレーションの特徴

シミュレータの設計として、我々は連続座標を採用した。座標空間の中でエージェントが行為し、その結果として環境及び環境内での自分の位置に変更を与え、行為の結果を環境から受け取る。ただし、学習にあたっては、学習時に用いる「状態」として、空間を離散なグリッドに区切ったものを用いている。

3.2 汎用性の確保

シミュレーションの汎用性という観点から、我々は、シミュレータを特定の問題に依存しない設計とし、さらに、カヌーレーシング問題においても、形や流速といった状況の異なる川に対応できるシミュレーションに変更した。本研究では、Jason と呼ばれる言語処理系と Java クラスで実現されているが、過去の我々の研究ではこの 2 つが独立性を保てない実現となっていた。そのため、クラス設計を全体的に見直し、再設計を行った。

具体的には、抽象クラスでフレームワークを作り、具体的な実装を subclasses で行うことで独立性を高め、川の形を柔軟に変更できるようにした。同様の方法を、環境を定義するクラスとエージェントを定義するクラスにも採用した。環境のみを取り替え、異なるテストベッドへの応用が可能となるよう、今後の展望をふまえた上でこれらの独立性も高めた。再度設計した具体的な Java クラスを下記に提示し、その概略を示す。

3.2.1 エージェントのみに関するクラス

AgentModel 主にエージェントの情報を取得する。 subclasses では、エージェントのスタート地点の座標の定義、及びエージェントによる行為選択を可能にした。

AgentView エージェントを定義し、描写する。

BasicAct 基本行為を定義する。動作後の詳細な座標は subclasses で定義することが可能である。

3.2.2 環境のみに関するクラス

RiverModel 川の定義とその情報を取得する。具体的には流速、向き、形を定義し、形の詳細は subclasses で定義することが可能である。また、ゴールの判定を行う。本研究では判定方法を変更し、川の形に応じて場合分けしている。

RiverView 川全般を描写する。

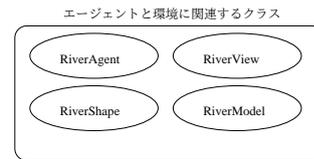


図 1: 設計変更前

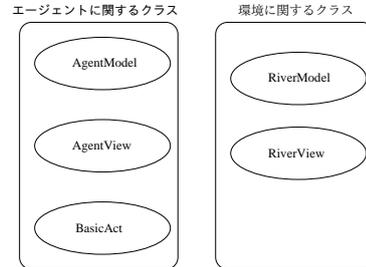


図 2: 設計変更後

3.3 学習方式の多様化

1 節で述べたように、実世界での行動のためには、問題に応じた柔軟な学習方式を取ることが必要である。そのため、学習方式もシミュレーションの問題設定とは独立に交換可能な設計とし、また、いくつかの学習方式を実装した。以下ではそれについて述べる。

3.3.1 Q と Sarsa

強化学習の学習方式に、Q 学習として知られている方策オフ型 TD 制御がある。単に状態行動対が全て更新され続けることさえ保証されれば、Q 学習は正しく収束する。そして、この条件と、ステップサイズ・パラメータに関する通常の確率近似の条件が保証される時、確率 1 で t 回目における行動価値関数は最適な行動価値関数に収束する [三上 00]。

一方、Q 学習の別策の一つとして挙げられるのが、Sarsa として知られる方策オン型 TD 制御である。これの特徴は、状態行動対の間の遷移に着目して、状態行動対に対する価値を学習する点である。この学習方式の収束性は、選択する行為と行動価値関数との関係に依存し、極限において行為がある状態における greedy な行為^{*2}に収束する場合、確率 1 で最適な行動価値関数に収束する [三上 00]。

3.3.2 ϵ -greedy と Softmax

greedy 手法は、greedy な行為を選択する最も単純な行動規則である。これに対し、大半は greedy な行為を選択するが、まれに小さい確率 ϵ で、行動価値推定量とは無関係に、一様に任意の行為を選択する手法を ϵ -greedy と呼ぶ。 ϵ -greedy の利点は、 t 回目のプレイで得られる各々の行為の行動価値推定量が、真の行動価値に収束することを保証している点である。これは、プレイ回数を増やし極限に至る時、全ての行為が無制限回試されることが理由である。

その他、Softmax と呼ばれる手法もある。推定価値を等級付けした関数によって、行為選択の確率を変化させるのが、これの特徴である。具体的には、 t 回目のプレイにおける行為を式 (1) ^{*3}で選択する。

$$\frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^n e^{Q_t(b)/\tau}} \quad (1)$$

*2 最も高いと推定された行動価値を持つ行為を指す。

*3 τ は温度と呼ばれる正定数である。

ϵ -greedy 手法では、行動を等しく選択するため、最悪であると考えられる行為を選択する可能性と、最適に近い行為を選択する可能性が同程度に高い。従って、選択肢の中で最悪である行為の程度が極めて悪い結果をもたらす場合、望まない結果を招くことになり得る。その点 Softmax 法では、greedy な行為に最も高い選択確率を与え差をつけるため、 ϵ -greedy で起こりうる欠点を解消することができる。

3.3.3 評価

我々はカヌーシミュレーションの例で、Q 学習と ϵ -greedy による学習を行い、得られる報酬と次なる状態を観測した。この時行為一回あたりに得られる報酬値を、下記の通りに設定した。

- その行為により、カヌーがゴールに到達する 100
- その行為により、カヌーが川の外にコースアウトする -100
- 上記の2点を除く、基本行為一回あたりの報酬 -1

エージェントが川の上端をスタートしてから、コース外に外れることなくゴールに到達するまで、あるいはゴール前に運悪く川の外に出てしまうまでを一つのエピソードとした。Q 学習では、エージェントがエピソードを 10000 回繰り返す頃には、学習前と比べて通算報酬^{*4}は 10 ポイントほど上昇する。最終的には、80 ポイントには満たず 75 から 76 ポイントに収束する。初めは川からコースアウトしていたが、学習を繰り返した後は確実にゴールできるようになった。

これらをふまえると、Q 学習によりエージェントの動作は明らかに改善が見られたが、一方で、通算報酬はある一定に収束していき限界が見られ、1 エピソード全体を通して必ずしも最適な行為選択をしているとは限らない。絶えず学習を行っているからであるが、エージェントが学習を行うプランと学習結果の利用のみを行うプランを状況によって選択することで、これを改善することは考えられる。

また、実世界では、たとえば「雨が降っている日に雨に濡れないよう目的地まで向かう」という意図に対して、「傘をさして歩く」という手段もあれば、「出発地から目的地までタクシーに乗る」など、一つの意図に対してその方法は様々である。

本研究は、エージェントが問題設定に合わせて自らプランを作成しそれを実行することを目標としており、強化学習の導入もその一環である。仮想世界のエージェントもまた、実世界同様複数のプランを考慮した上で、状況に合わせて選択できることが望ましい。その場合、選んだプランによって、異なる学習方式による行為をとる方がよいことも考えられるであろう。そこで我々のシミュレータは、学習方式を切り替えることによって、エージェントが様々な学習方式の選択肢を取り入れることを可能としている。

4. 考察

今回各クラスの役割を細かく分離して再設計したことは、川の定義の変更を容易にした。異なる形や流速の川においてシミュレートする際は、RiverModel の子クラスを取り替えるだけで、自由に様々な形の川のシミュレーションができる。しかしながら、変更できる川の形式にはまだまだ制限がある。画面上において上から下にかけて下以外は対応できず、U 字型や渦を巻くように下っていくような、複雑な形式のカヌーレー

シングはシミュレートすることができない。これを解決するには、一つにゴールの判定方法の変更がある。

現状の判定方法では、ある一定の方向からゴールラインを越えたか否か、すなわちエージェントの y 座標を、アプリケーションの画面上におけるゴールラインの y 座標と比較することで判定している。従って、川上であるか、それとも川下であるのかというのは、あくまで位置的な比較のみである。だが実際には、実世界は平面ではなく高さという概念があり、川もまた z 軸方向の高度をふまえた上で、川上と川下が決まる。

今後環境の複雑化や異なるテストベッドに対応させていくには、仮想世界に高さの概念を取り入れることは勿論であるが、例えば風や障害物など、実世界で考え得る様々な外乱の影響も考慮する必要がある。そしてそれを情報として定義し、エージェントが学習に活用させることが今後の課題である。

5. 終わりに

本論文では、汎用性の向上と学習方式の多様化に焦点を当て、連続的な仮想世界におけるシミュレーションの実装について述べた。環境に依存したクラス設計を改善したことで、エージェントの定義はそのままに、カヌーレーシングとは条件の異なる例題の取り入れに向けて一歩前進した。

エージェントは、あくまでアプリケーション上の座標を認識して移動する。それは川の形式が異なっても、あるいは川ではなくとも同様である。現段階では、目標達成の判定を、川の下端に到達したか否かで行っている。従って、現状定義している川のようにゴールラインが直線であれば、環境の定義を川ではなく平地に、あるいは移動手段をカヌーではなく徒歩を想定したものに变更したとしても、エージェントは何ら影響を受けず、問題ないと結論づけられる。

また、本研究ではエージェントによる問題解決のためのプランの作成を可能にすることを目標とし、学習の多様化も進めて来た。学習方式として Sarsa、行動価値手法として Softmax 法を新たに取り入れ、カヌーレーシングの川下りは、Q 学習や ϵ -greedy による選択とは異なる結果となった。学習方式をはじめ、まだ導入していない学習方法も今後取り入れることで、エージェントのプラン作成に活用したい。

各々の学習方式や行動価値手法にはメリットやデメリットがある。それらは一概に優劣をつけることは難しく、実際には、状況に応じた利用が望ましい。エージェントのプラン作成にあたり、強化学習によって得られる手段が複数あり、それらが異なるものである時、エージェントがいかにして評価し選択するのか、その基準や評価方法も明確にしていく必要がある。そして、基本行為の詳細化など環境を実世界に近づけていくことを今後の課題としたい。

参考文献

- [高田 12] 高田 司郎, 新出 尚之: 行為のアトラクター状態を考慮した知能ロボットについて, in *Proc. of JAWS2012* (2012)
- [高田 13] 高田 司郎, 新出 尚之, 濱砂 幸裕, 波部 斉, 藤田 恵: アトラクター状態を用いた実世界における基本行為の学習について, 情報処理学会研究報告 2013-MPS-92, No. 24, pp. 1-6 (2013)
- [三上 00] 三上 貞芳, 皆川 雅章: 強化学習, 森北出版株式会社 (2000)
- [辰野 05] 辰野 勇: カヌー&カヤック入門, 山と溪谷社 (2005)

*4 1 エピソードにおける報酬値の合計