平成28年度　卒業論文

流れを考慮した学習を行える
川下りシミュレーション環境の構築

平成29 年2 月13 日

奈良女子大学　理学部　情報科学科　新出研究室
学籍番号　13253085

胡好皓

概要

人のように考えて行動できるようなロボットが望まれている。いきなり現実世界に実装することは困難であるため、初期段階として現実世界を想定した仮想世界上にロボットがカヌーを操縦し、川下りを行うシミュレータが先行研究において実装されている。しかし、現実の川は流れの速度や向きが変化するが、従来のシミュレータはこれを考慮しておらず、漕ぎ方の学習も学習後の行動も一定の流れのもとで行っている問題があった。そこでシミュレータに川の流れの変化を導入し、学習方法を改善して実験を行った。

目次

目 次

1　はじめに

近年、世間では AI が注目されるようになり、人のように自身の経験から得た知識と周りの情報から物事を判断して行動できるようなロボットが期待されている。我々の研究室はこのような自律型ロボットの実現を最終目標としている。しかし、現実世界はランダム性を備えた環境であり、現実世界で適した行動をいきなりロボットに獲得させるのは困難である。そのため、初期段階として実世界を想定した仮想世界上にシミュレータを作成する必要がある。その一環としてカヌーレーシングを題材とし、先行研究においてロボットをカヌーに乗せ、ロボットがカヌーを操縦し、川下りを行うシミュレータが実装されている [1, 2, 3, 4]。

これまでのシミュレータは川の流れを最初に与えた状態から変化しないと仮定し、強化学習を行うにあたっても川の流れを全エピソードで一定と仮定して学習を行っていた [4]。しかし、現実の川の流れは刻一刻と変化するため、川の流れの変化をシミュレータに導入した。学習も、川の流れの変化をも考慮した学習に改善し、実験を行った。

2　自律エージェント

2.1　自律エージェントとは

自律エージェントとは、自ら目標を持ち、目標を達成しようと方法を考え、方法を選択して実行し、問題を解決しようとするシステムである。カヌーシミュレータでは人のように考えて行動できるような知能ロボットが、周囲の環境を考慮しながら、カヌーを操縦し、より速くゴール地点に到達することを想定している。このロボットは自律エージェントであり、BDI モデルをもとに実装している。BDI モデルについては以下に述べる。

2.2　BDI モデル

BDI モデルは、人が目標達成に向けて行動する一連の動きを再現したエージェントのモデルの一つである。「信念 (Belief)」、「願望 (Desire)]、「意図 (Intention)」の概念を用いて、行動選択を行う。信念とはエージェントが環境に関して持っている情報、願望とは実現することが望ましいと考えられる事柄である。その中で現在の信念に照らし、どれを最終的に実現したいかを選んで目的とする。目的を達成するための手段を選択し、それに専念すると決意したものが意図である。願望も目標も複数同時に存在する場合がある。

3　カヌーシミュレータ

3.1　機能

先行研究で実装されたカヌーシミュレータは、現実世界を想定した仮想世界上にて、ロボットをカヌーに乗せ、ロボットがカヌーを操縦し、川下りを行うシミュレータである。スタート地点からゴール地点までを川岸や障害物に衝突しないよう、カヌーの安定状態を保ちながら、より少ない行動回数でゴールに辿り着くことを目指すものである。

初めにスタート地点及びゴール地点、川の形状等の初期条件を与えてからシミュレータを実行すると、シミュレータはカヌーがより速くゴールに辿り着ける道順を、複数回川下りを行わせることで探索し、その時の最適な行動を学習する。川下りの際のロボットの行動はカヌーの操縦を川の流れに任せる、前進させる・後退させる・左折させる・右折させる、の 5 種類から選択される。最後に、学習結果から最適と判断された漕ぎ方による川下りの様子が表示される。

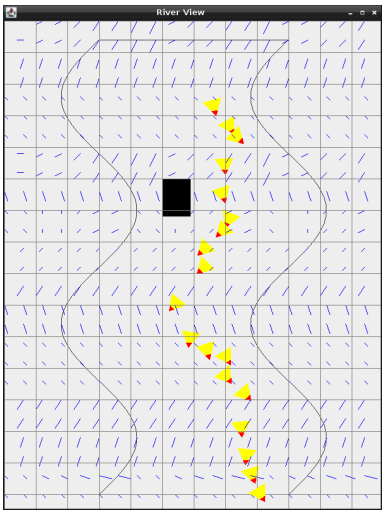

図 1: カヌーシミュレータの実行例

図 2: カヌーの図

カヌーシミュレータの画面表示を図 1 に示す。空間の大きさや川の形は任意で指定でき、川の流れは青色の線で表示されている。空間の座標は連続で与えることができる。川の流れに関しては、強化学習において状態を有限通りとするために空間を任意の大きさに区切り、その中は川の流れが等しいものとしている。カヌーは図 2 のように黄色の三角形で表現しており、その一角の赤色の三角形がカヌーの先頭方向を表している。スタート地点は図での川の上端から選ぶ。ゴールは川の下端とする。下端の任意の地点に到達すればゴールとする。

3.2 改善点

3.2.1 改善点 1　川の流れのランダム化

現在の実験では川を 12 × 16 のグリッドで分割している。これまでの実験環境は川の流れを最初に与えた状態から変化しないと仮定してきたため、従来のシミュレータはその 1 グリッド毎に川の流れのベクトルの x 成分、y 成分を固定値で設定していた。実験環境の川を現実に近づけられるよう、以下の式で川の流れを設定し、学習時や実験時に 1 エピソード毎に随意に変更できるようにした。式中の a、b、k は適当な定数であり、今回の実験では $a = 5$、$b = 2$、$k = 3$ としている。

$$x \text{ 成分} : 0 + (0 \sim 1 \text{ の乱数} - 0.5) \times a \tag{1}$$

$$y \text{ 成分} : k + (0 \sim 1 \text{ の乱数} - 0.5) \times b \tag{2}$$

今回の実験では、1 エピソード中にカヌーが川の中の同じ地点を 2 度通るということが起こらないため、カヌーが 1 回漕ぐたびに流れを変更したとしても、1 エピソード毎に変更するのと原理的に変わらない。そのため、流れの変更は 1 エピソード毎とする。

3.2.2 改善点 2　基本行為

カヌーの基本行為の中でも左折及び右折は、従来はカヌーの向きと関係なく、常にシミュレーション画面での真下の向き (0 °) から 45 °回転するように設定されていたが、直前のカヌーの向きをもとに回転するように変更した。これは、卒業研究発表後に判明した問題点の改善である。なお、第 4 節の実験 1 と 2 はこの改善より以前に行ったものである。

4　強化学習

4.1　強化学習

強化学習 [5] とは、ある環境状態 s の下で、行動 a を選択する価値 $Q(s,a)$ を学習する方法である。ある状態 s のとき、価値 $Q(s,a)$ が最も高い行動 a を最適な行動とする。学習中は状態 s と行動 a の組み合わせを色々試し、試行錯誤の積み重ねにより、それぞれの状態での各行動の価値を推定する。

今回のシミュレーション実験では、学習方法には Q 学習を用い、学習中の行動選択は ϵ-greedy 法を用いた。それらについては以下に述べる。

4.2　Q 学習

Q 学習 [5] は学習方式の一つで、状態行動価値の推定を即時報酬と遷移後の状態の推定価値を用いて求める手法である。ある時刻 t での状態 s_t から次の時刻 $t+1$ での状態 s_{t+1} へ遷移したときの更新式は以下のとおりである。γ は割引率といい、遠い将来に得られる報酬ほど、割り引いて評価するためのものである。α はステップサイズ・パラメータである。また、a_t は状態 s_t で取った行動、r_{t+1} はその行動によって時刻 $t+1$ で得た即時報酬を表している。

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)) \quad (3)$$

4.3　ϵ-greedy 法

ϵ-greedy 法 [5] は一定の確率 ϵ で取り得る行動のうち一つをランダムに選び、$1-\epsilon$ の確率で最大の価値を持つ行動を選択する方法である。

4.4　改善点

先行研究では強化学習の状態を川の位置のみに限定していた。しかし、現実世界では川の流れもカヌー操縦に影響を与えるため、川の位置とともに川の速度も学習時に考慮する状態として強化学習に取り入れた。

5　実験

実験１

川の位置のみで学習する従来の学習方法で学習したシミュレータと川の位置と流れを両方考慮して学習する現在の学習方法で学習したシミュレータに、1 エピソード毎に川の流れを変更しながら、どちらにも 3 万エピソードをそれぞれ学習させた。その後、流れ方が異なる 1 千パターンの川で川下りの結果を比較した。この 1 千パターンの川の流れパターンは双方のシミュレータで同じものを用いるが、強化学習時のものとは異なる組み合わせになるようにしている。3 万エピソードの学習と 1 千パターンの川下りを 1 セットとし、各シミュレータで 10 セット実行し、ゴールに辿り着けた回数とそのときの行動回数を比較した。

学習で使用するパラメータは $\epsilon = 0.05$、$\gamma = 0.9$、$\alpha = 0.1$ としている。実験結果を下表にまとめた。

学習方法	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
川の位置のみ	703.4	13.3
川の位置と速度	643.3	14.6

表 1: 実験 1 の結果

表 1 より、ゴールに辿り着いた回数の平均やそのときの行動回数の平均は予想と反し、川の位置のみで学習を行ったシミュレータの方が川の位置と速度で学習したシミュレータよりも良い結果を返していることがわかる。

実験 2

実験 1 から従来の方法で学習を行ったシミュレータの方が現行の方法で学習したシミュレータよりも良い結果を返していることがわかったが、その理由として、現行の方法では学習内容が従来よりも増加したゆえ、学習回数が足りず、良い結果を返すことができていないのではないかという仮説を確かめるため、実験 1 と同じ実験環境下で、学習回数のみを 1 回から 5 万回まで 1 万回刻みに変更して実験を行った。結果を表 2、表 3 及び図 3、図 4 にまとめた。

学習回数 (万回)	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
1	687.7	13.5
2	702.6	13.9
3	703.4	13.3
4	704.4	13.5
5	706.2	13.5

表 2: 川の位置のみで学習した場合の結果

学習回数 (万回)	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
1	627.6	14.4
2	626.4	14.7
3	643.3	14.6
4	664.5	14.4
5	629.9	14.5

表 3: 川の位置と速度で学習した場合の結果

実験の結果、現行の方法と従来の方法どちらも 1 万回程度の学習で学習結果は安定しており、学習回数の不足が原因ではないと考えられる。

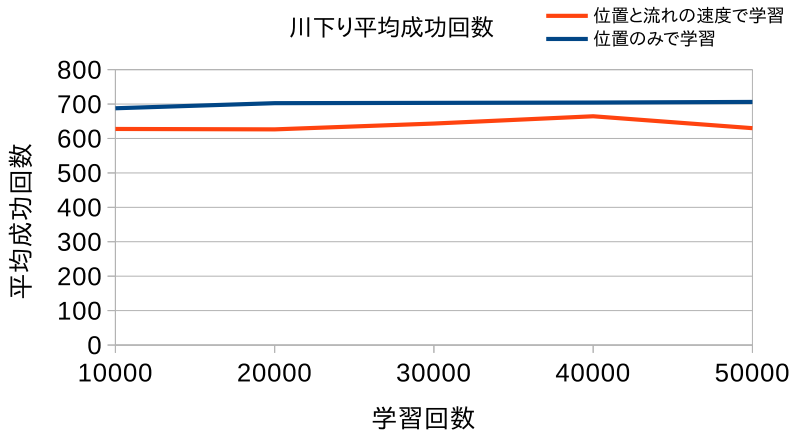

図 3: 川下りの平均成功回数の比較

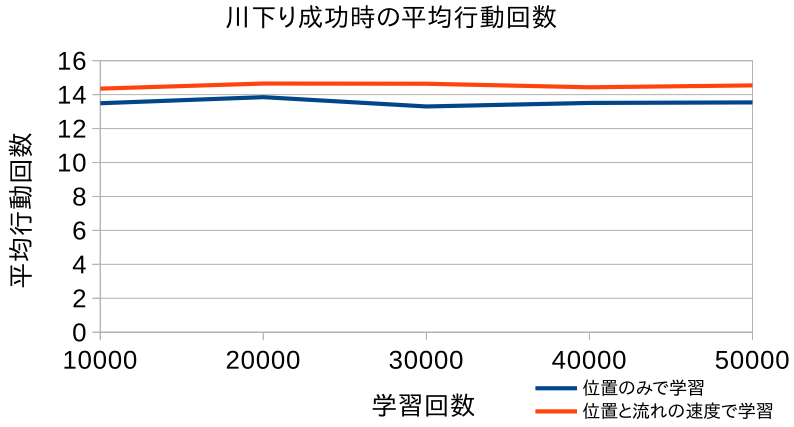

図 4: 川下り成功時の平均行動回数の比較

実験 3

実験 2 の後にプログラム全体を見直した結果、3.2.2 に述べたようにカヌーの基本行動に物理的な法則に即していない部分を発見した。その部分を改良し、実験 1 と同じ条件下で再度実験を行った。結果を表 4 にまとめた。

この場合も実験 1 と同様、川の流れを考慮しない学習の方が成績がよいという結果になった。

学習方法	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
川の位置のみ	754.7	14.9
川の位置と速度	688.7	16.0

表 4: 実験 3 の結果

実験 4

実験 2 と同様な実験を基本行動改良後のシミュレータでも行った。結果は表 5、表 6、図 5、図 6 の通りである。

学習回数 (万回)	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
1	754.7	14.6
2	744.6	14.5
3	754.7	14.9
4	746.5	14.1
5	720.4	13.5

表 5: 川の位置のみで学習した場合の結果

学習回数 (万回)	ゴールに辿り着けた平均回数 (回)	ゴールに辿り着くまでの行動回数の平均 (回)
1	698.3	15.8
2	706.4	15.9
3	688.7	16.0
4	705.3	16.0
5	715.5	16.2

表 6: 川の位置と速度で学習した場合の結果

この場合も実験 2 と同様、1 万回程度の学習で学習結果はほぼ安定しており、学習回数の不足ではないと考えられる。

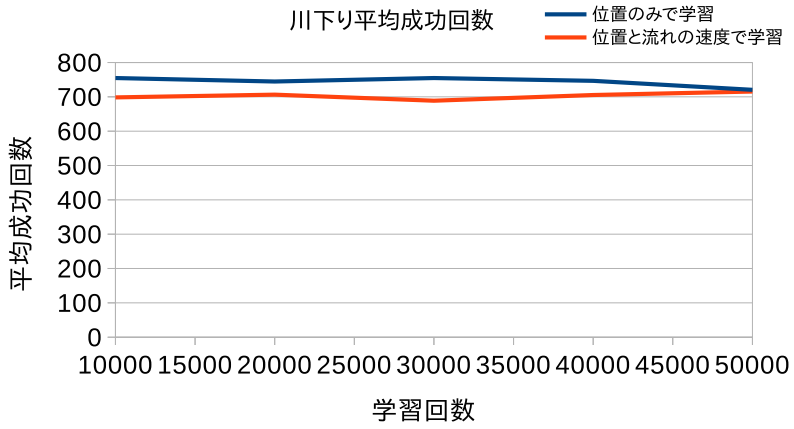

図 5: 川下りの平均成功回数の比較

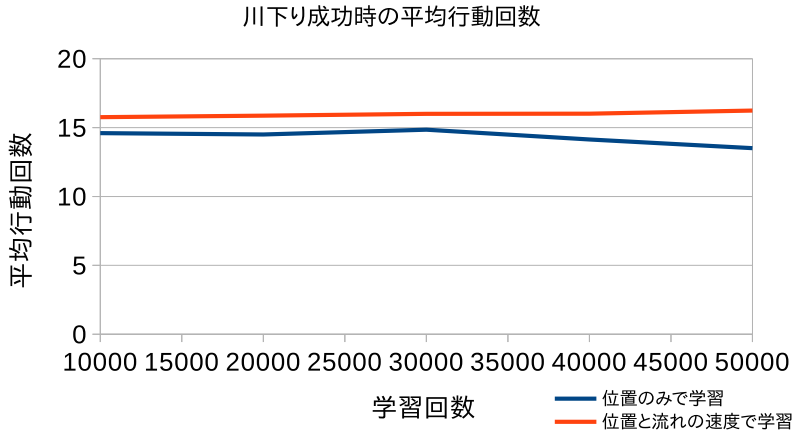

図 6: 川下り成功時の平均行動回数の比較

考察

実験より、事前の予想に反して、の川の位置のみで学習を行ったシミュレータの方が川の位置と速度で学習したシミュレータよりも良い成績となるという結果を得た。その理由としては、以下のようなことが考えられる。まず、川の位置のみで学習した場合は、学習時には 1 エピソード毎に川の流れを変化させているため、同じ漕ぎ方でも場合によって漕いだ結果のばらつきが大きいこと、すなわち、川の流れが不安定であるという傾向を学習する。一方で、現在の実験では、川の形状があまり屈折していないことや、障害物があ

まり大きくなく数も少ないこと、位置も川の端寄りであることなどから、おおむね川の中央付近をたどれば容易にゴールまで到達できるような問題設定となっている。その結果、流れの不安定さに対抗して川の中心付近にできるだけ寄ろうとするような漕ぎ方を学習することで、結果的にゴールに辿り着く確率が上がり、川下りの成功率が高くなったのではないかと考えられる。一方、流れを考慮した学習で成功率が低い理由としては、流れを考慮したため川幅をいっぱいに使うような漕ぎ方を学習し、その結果川の外に出る危険性がかえって高まっている可能性が考えられる。

また、ゴールに辿り着くまでの行動回数の平均も、川の真ん中よりを漕ぐことで、比較的まっすぐな漕ぎ方となって近道になるため、少ない行動回数でゴールに辿り着けたのではないかと考えられる。

6　まとめ

本研究ではシミュレータの学習方法を、川の位置と流れの速度を考慮した学習に改良したことにより、シミュレータは川の流れが変化しても対応できるようになった。また、実験を通し、川の流れを考慮した学習を行っても、有利になるとは限らないという結果を得た。

今後の課題を述べる。今回の実験では、流れを考慮しない学習の方がよい成績となったが、これについては、第 5 節で述べたように、現在の実験では、おおむね川の中央付近をたどれば容易にゴールまで到達できるような問題設定となっているためではないかと考えられる。そのため、流れを考慮しない学習の方が本当によいかどうかを検討するには、今後の課題として問題設定を変更する必要がある。例えば、現在の実験では川の形状があまり屈折していないため、湾曲を増やし、まっすぐにはゴールに辿り着きにくいような形にする。3.2.1 の定数 a,b,k を変更し、川の流れをより急激にする。障害物の数を増やし、形状を変更することがあげられる。これらの変更によって、流れを考慮しないとぶつからずに漕ぎにくいような問題設定になるのではないかと考えられる。一方で、流れの変更に当たっては隣同士のグリッドで流れは極端に変化しないよう、考慮することも自然界に近づけるためには必要である。

学習方法に関しても、Q 学習を Sarsa[5] に変更してみるとよいのではないかと考えられる。Sarsa の方が、端からはみ出にくいような漕ぎ方を学習するのに適しているため、川の流れを考慮する場合に有利になる可能性がある。また、現実に近づけたシミュレータの作成も今後の課題の一つである。

謝辞

本研究を行うにあたり、平日休日を問わず夜遅くまで長い時間にわたり熱心にご指導、ご助言くださった新出尚之准教授に深く感謝し、厚く御礼を申し上げます。また、新出研究室の皆様に深く感謝の意を表します。

参考文献

[1] 諏佐歩. 行為のアトラクター状態を考慮したカヌーシミュレータ環境の構築. 卒業論文, 奈良女子大学理学部情報科学科, 2016.

[2] 亀村美佳. 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. 修士論文, 奈良女子大学大学院人間文化研究科, 2016.

[3] 柚木静香. 実世界のエージェント構築のためのシミュレーション環境の実現. 卒業論文, 奈良女子大学理学部情報科学科, 2015.

[4] 宮田怜奈. カヌーレーシングのシミュレーション環境における強化学習について. 卒業論文, 奈良女子大学理学部情報科学科, 2015.

[5] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning.* The MIT Press, 1998. (三上貞芳 and 皆川雅章 共訳. 強化学習, 森北出版株式会社, 2000).