7W-05

地形を考慮したValue Iteration Networksによる経路探索

工藤純暉　長名優子

東京工科大学 コンピュータサイエンス学部

1　はじめに

経路探索に関する研究は古くから盛んに行われており、多くのアルゴリズムが提案されている。近年では、深層学習と強化学習とを組み合わせた深層強化学習の分野においても経路探索に関する研究が行われている。そのような手法の1つとして、Deep Q-Network[1] を用いた経路探索が提案されている。Deep Q-Networkでは、ある状態のときにどの行動をとるかを試行錯誤によって学習することができる。しかし、計画をたてて行動することはできないため、環境の変化に弱く、別の環境で学習した情報を別の環境で利用することは難しかったり、フィールドのサイズが大きくなると学習が難しくなるなどの問題点がある。それに対し、Value Iteration Networks[2] という手法が提案されている。この手法では、プランニング自体を畳み込みニューラルネットワークを用いて学習することで、大きなフィールドにも対応することができ、他の環境で学習した情報を利用して未知の環境でもゴールにたどりつくことが可能となっている。また、ゴールが複数存在するような環境や複数のエージェントが存在するような状況にも対応することのできるValue Iteration Networksを用いた経路探索 [3] も提案されている。この手法では、複数の目的地が存在する場合には、すべての目的地を考慮したうえで価値マップを生成することで対応している。また、複数のエージェントが存在する場合には、他のエージェントがいることで発生する混雑状況を考慮して行動選択を行うことで対応している。

本研究では、Value Iteration Networksを用いた経路探索において、地形の影響を考慮したが行える方法を提案する。

2　Value Iteration Networks

Value Iteration Networks[2] は畳み込みニューラルネットワーク [4] による報酬の推定、VI (Value Iteration) モジュールによる報酬マップからの価値マップの生成、注意、行動の選択の4つの処理から構成されている (図1)。

2.1　畳み込みニューラルネットワークによる報酬と状態遷移の推定

まず始めに、畳み込みニューラルネットワークを用いて報酬 R の推定を行う。畳み込みニューラルネットワークへの入力はエージェントの観測である。観測は状態から得られる情報であるが、エージェントが状態の情報をすべて観測できるとは限らないため、状態と観測が一致しないこともある。Value Iteration Networkで経路探索問題を扱う場合には、観測としてフィールドのマップデータとエージェントの座標を用いる。畳み込みニューラルネットワークでは、報酬や状態遷移を出力するように学習を行う。マップの各座標における報酬を表したものを報酬マップと呼ぶ。

2.2　報酬マップからの価値マップの生成 (VIモジュール)

次にVIモジュールにおいて報酬マップからの価値マップの生成が行われる。VIモジュールでは、報酬 $\overline{R}$ に対して畳み込み演算を行い、各行動に対する行動価値 $\overline{Q}$ を計算する。行動価値 $\overline{Q}$ はエージェントが行動 a を選択したときに、行動 a に対して現時点での報酬と合わせてどのくらいの報酬になっているかを計算することで得られる。次に最もよかった行動に関する価値 $\overline{V}$ を出力する。出力された価値 $\overline{V}$ をもう一度報酬 $\overline{R}$ とともに入力とし、これらの処理を繰り返すことで報酬を伝播することができる。

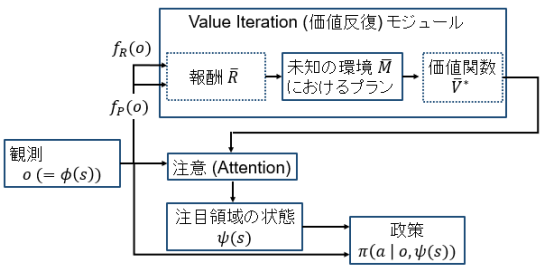

図 1: Value Iteration Networksの処理の流れ

Route Search using Value Iteration Networks considering Topography
Junki Kudo and Osana Yuko(Tokyo University of Technology, osana@stf.teu.ac.jp)

2.3 注意・行動の選択

注意モジュールでは、観測から注目領域の切り出しを行う。経路探索問題では、エージェントの位置の周囲のみに着目する。行動選択は、注目領域の価値マップの値に基づいて softmax 関数により行われ、8 近傍のマスのうち価値が高いマスに移動するような行動が選択される。

3 地形を考慮した Value Iteration Networks による経路探索

提案する地形を考慮した Value Iteration Networks による経路探索では、各マスの高さを表す地形情報は地形マップとして取得可能であることを前提としている。行動選択をする際に、上りであれば平地や下りに比べて移動に余計に時間がかかることを考慮して、価値の修正を行い、地形を考慮した行動選択が行えるようにする。

3.1 対象とする課題

エージェントが動き回るフィールドとしては、格子状に区切られたものを想定し、1 つのマスには 1 つのエージェントのみが存在できるものとする。フィールドには、出口が設置されており、エージェントは出口を目指して移動することになる。各試行の開始時には、エージェントは障害物と重ならないような位置にランダムに配置される。

エージェントはフィールド全体を観測できるものとし、地形 (高さ) の情報を含むフィールド全体の情報とエージェントの位置を観測として使用する。エージェントは 8 近傍のいずれかのマスに移動するような行動をとる。エージェントは Value Iteration Networks から得られた価値マップに基づいて、自分の 8 近傍のマスのぞれぞれの価値を取得する。また、地形情報を参考にし、各マスの価値を再計算が行われる。その後、8 近傍のマスのうち、価値が一番大きいところに移動する。ただし、移動先に障害物が存在する場合にはそのマスを移動可能なマスから削除し、残ったマスの中から改めて価値が一番大きいマスに移動するものとする。

3.2 地形の高低差による影響の計算

マスの高低差を考慮した経路探索を実現させるのにあたり、経路の高低差による価値の修正を行う。

地形を考慮したマス i における行動 a の価値 $V'_{i(a)}$ は

$$V'_{i(a)} = V_i - \beta \cdot J_{i(a)} \tag{1}$$

で与えられる。ここで、V_i はマス i の状態価値、β は地形の影響を決める係数である。また、$J_{i(a)}$ はマス i における行動 a に関する正規化した高低差の影響値であり

$$J_{i(a)} = \frac{\exp(j_{i(a)})}{\sum_{a'=1}^{N} \exp(j_{i(a')})} \tag{2}$$

で与えられる。ここで、N は移動可能なマスの数である。$j_{i(a)}$ はマス i における行動 a に関する高低差の影響値で

$$j_{i(a)} = \sum_{j \in S_a} h_{j(a)} \tag{3}$$

で与えられる。ここで、S_a は行動 a をとったときの経路上のマスの集合、$h_{j(a)}$ はマス j で行動 a をとったときの高低差の影響である。$h_{j(a)}$ はマス j と隣接するマスとの高低差によって決まる。マス j から行動 a をとったときに移動する先のマスがマス j よりも高い位置にあればその経路は上りであり、移動に時間がかかると考えられるため、$h_{j(a)}$ の値は正の値に設定される。それ以外の場合は、$h_{j(a)}$ は 0 とする。

4 計算機実験

計算機実験を行い、提案手法において高低差を考慮した経路探索が実現できることを確認した。

参考文献

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. Grave, I. Antonoglou, D. Wierstra and M. Riedmiller : “Playing Atari with deep reinforcement learning,” NIPS Deep Learning Workshop, 2013.

[2] A. Tamar Y. Wu, G. Thomas, S. Levine and P. Abbeel : “Value iteration networks,” Annual Conference on Neural Information Processing Systems, 2016.

[3] 蔡金雨, 李子龍, 長名優子 : “複数の目的地・エージェントを考慮した Value Iteration Networks による経路探索.” 情報処理学会第 83 回大会, 2021.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis : “Human-level control through deep reinforcement learning,” Nature, No.518, pp.529–533, 2015.