F-038

強化学習を用いた Dynamic Window Approach（DWA）パラメータの最適化手法
Dynamic Window Approach (DWA) Parameter Optimization Using Reinforcement Learning

今村 心祐† 中村 潤†
Shinyu Imamura Jun Nakamura

1. はじめに

近年, 自動運転技術の研究開発が国内外で盛んに進められている自動運転を実現させるためは, 車両が安全に走行するために, 路上という動的な環境下での経路計画が必要である. そのような経路計画アルゴリズムの一つとして, Fox らが提案した, Dynamic Window Approach（DWA）が挙げられる(Fox, et al. 1997). DWA では, 設定された時間おきに, 車両の予測軌道をいくつか求め, それぞれの軌道での並進速度（v）, 角速度（ω）を, 目標地点への角度の差分（heading）, 障害物との距離（distance）, 並進速度（velocity）とそれぞれの重みパラメータ（α, β, γ）によって構成された評価関数に代入し, 最も評価が高かった予測軌道を車両がゴールにたどり着くまでトレースし続ける. このように, DWA は一定時間おきに経路を計算するという特性があるため, 動的な環境下での経路計画アルゴリズムとして有効であると考えられる. しかし, DWA では, 決められた制限時間までにゴールにたどり着くことが不可能な状態である局所解に陥ってしまうことがあることが確認されている. 本研究では, 袋小路型の障害物などの周囲環境に起因する局所解と, 重みパラメータの調整不良に起因する局所解の二つを解決されるべき課題として定義した. 本研究では, 一定時間おきに, 周囲環境と車両の状態から, 重みパラメータを動的に調整する手法を提案する. ここで, 重みパラメータを動的に求める方法として, Dobrevski らの先行研究(Dobrevski and Skočaj, 2020)を参考に, DWA の評価関数における 3 つのコストを, ニューラルネットワークで処理し, 重みパラメータを決定する手法を採用する. 本研究では, 先行実験とは異なり, 強化学習の観測値として, DWA における heading, distance, velocity にそれぞれの重みパラメータを乗算した結果である 3 つのコストを用いることにより, 強化学習の簡素化がなされている.

本稿では, 最初に提案手法の説明を行い, 次にシミュレーション環境の紹介を行う. そして, 実験手法及び評価方法を紹介し, 予備実験を紹介する. おわりに, 今後の課題とまとめを述べる.

2. 強化学習型 DWA プログラムの仕様

2.1 DWA プログラムの制御フロー

本研究で実装している DWA は, プログラムを簡略化するために, 評価関数の実装を変更している. 一般的な DWA では, 評価関数で最も評価が高い値が得られた軌道を最適な軌道として出力するが, 本研究の DWA では, 軌道ごとのコストを計算し, 最もコストが小さかった軌道を最適な軌道として出力する. まず, ロボットカーに定義されたスペックと, 軌道計算時のロボットカーの状態から, 次の軌道計算時までに出力可能な並進速度, 角速度のペア（v, ω）をプログラムに定義された解像度分で計算する. 速度の最大値から計算された並進速度（v）を減算した値を速度のコストとし, 目標地点と計算された角速度（ω）でのロボットカーの角度との誤差を角度のコストとする. そして, 障害物との距離の逆数を障害物のコストとし, 得られた 3 つのコストにそれぞれのコストに, 定義されたそれぞれの重みパラメータを乗算し, 最もコストが小さかった軌道を最適な軌道として出力し, ゴールに到達するまで上記の計算を繰り返し, 経路計画を行う. なお, 本研究において, 重みパラメータは[0, 1]の範囲における連続値とする.

2.2 強化学習の設定

本研究では, OpenAI gym に準拠した自作のシミュレーション環境を構築し, シミュレーションを行う. また, 強化学習については, Stable Baselines3 の PPO プリセットを利用し, 学習を行う. 学習エージェントに観測値として与えられるのは, DWA による軌道計算時に最適な軌道として出力された軌道の速度のコスト, 角度のコスト, 障害物のコストとし, 方策として速度の重みパラメータ, 角度の重みパラメータを出力する.

2.3 報酬設計

学習時の報酬（R）は Dobrevski らの先行研究を参考に, 以下の式 1 のように定義する. なお, 式の定義は次のとおりである.

d_g: ロボットとゴールの距離

$$R = \begin{cases} 100 & , d_g < 1.0 \\ -30 & , 500 \text{ 回以上軌道計算を行った場合} \\ 0 & , \text{上記以外の場合} \end{cases}$$

式 1 報酬設計

なお, 学習における割引率は次のとおりである.
γ: 0.99

3. シミュレーション環境

シミュレーション環境は二次元空間で表され, ゴールの座標は（10, 10）で固定し, ロボットの初期座標は（0, 0）で固定とする. 障害物は袋小路であり, 横の長さは[3, 6], 縦の長さは[2, 6]の範囲でランダムに決定され, エピソードごとに変化する. これによって様々な形状の袋小路を再現する. シミュレーション画面のイメージ図は図 1 のとおりである.

† 中央大学 Chuo University

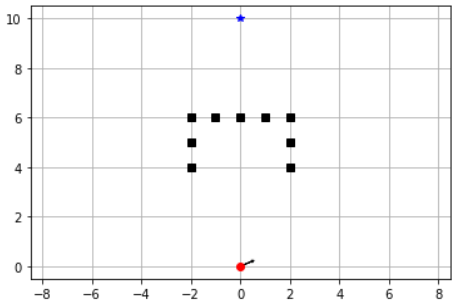

図 1　シミュレーション画面のイメージ図（袋小路が横 5 縦 3 の長さの形状の場合）

4.　実験手法

本研究では, 500 回の軌道計算の実行もしくは, ゴールまでの距離の数値が 1 となった場合をエピソードの終了判定とし, 100 エピソード経過時点での学習モデルを搭載したロボットカーで 100 エピソードを新たに走行させる. そして, 1 エピソードで重みパラメータを変更しない DWA を搭載しているロボットカーと, 軌道計算実行時にランダムで重みパラメータを変更する DWA を搭載したロボットカーをそれぞれ 100 エピソード走行させる. なお, 走行はシミュレーション環境上で実施される.

5.　評価方法

本研究では, それぞれの手法における DWA を搭載したロボットカーが, 100 エピソードあたりのゴールした回数の割合によって比較する. そして, ゴールした割合をそれぞれの手法における局所解への陥りにくさと定義し, 評価する.

6.　予備実験と考察

評価関数のパラメータの動的な変更によって, ゴールする割合への影響を調べるための予備実験として, 以下の条件で 2 種類のロボットカーA・B の走行実験それぞれ行い, ゴールした割合を比較した.

6.1　ロボットカーA の実験

ロボットカーA は, DWA の評価関数の速度のコストと角度のコストをエピソードごとに[0, 1]の範囲でランダムに設定し, 走行実験を行った. 環境は, ゴールの座標（10, 10）で固定し, ロボットの初期座標は（0,　0）で固定した. なお, 障害物は本研究とは異なり, 図 1 の障害物（黒い四角）の座標で固定した. なお, エピソード数は 100 として, 結果を外部のファイルに記録した.

6.2　ロボットカーB の実験

ロボットカーB は, DWA の評価関数の速度のコストと角度のコストを軌道計算ごとに[0, 1]の範囲でランダムに設定し, 走行実験を行った. 環境はロボットカーA の実験と同様に図 1 のように固定した. なお, エピソード数は 100 として, 結果を外部のファイルに記録した.

6.3　結果

表 1 は, 100 エピソードにおけるロボットカーごとのゴールした割合を表している. なお, ゴール割合を求める式は式 2 と定義する.

ゴール割合[%] ＝ ゴールした回数 ÷ 総ステージ数× 100

式 2　ゴール割合

表 1　予備実験のゴール割合

	ロボットカーA	ロボットカーB
ゴール割合	6.00%	47.00%

6.4　考察

表 1 から, エピソードごとに[0, 1]の範囲でランダムに設定したロボットカーA のゴール割合は 6.00%であり, 軌道計算ごとに[0, 1]の範囲でランダムに設定したロボットカーB のゴール割合は 47.00%である. 評価関数の重みパラメータを動的に変更することによって, 局所解に陥りにくくなることが確認された. つまり, パラメータを動的に変更する際に, ロボットカーB の実験のようにランダムで変更するのではなく, 強化学習によって最適化を行うことによって, 周囲の環境が変化した際においても, ゴール割合に向上の余地があることが予測される.

7.　今後の課題

現状では, 強化学習プログラムは試作段階である. 主な課題は, 強化学習における出力の範囲が, 性能へ大きく影響しており, 慎重な調整が必要なことである. 例えば, 前述のとは別の予備実験において, ニューラルネットワークから出力される速度のパラメータを[0, 1.0]とした場合, 出力値が連続して 0 となることが確認され, ロボットカーがその場で止まってしまう局所解に陥ることが観測された.

今後は, 強化学習プログラムの実装の調整を行い, 前述の実験手法, 評価方法にて提案手法の有用性を判断する.

8.　まとめ

本研究では, DWA における袋小路型の障害物などの周囲環境に起因する局所解と, 重みパラメータの調整不良に起因する局所解の二つを課題として定義している. 解決手段として, 強化学習を用いて, 通常は固定されている重みパラメータを動的に最適化する手法を提案する. Dobrevski らの先行研究と大きく異なる点は, 強化学習の際の観測値として, 軌道計算時の heading, distance, velocity の三つの評価値を用い, heading, velocity の重みパラメータを出力する手法を提案している点である. 今後は, 強化学習プログラムを実装し, 提案手法の有効性を評価する予定である.

参考文献

[1] D. Fox, W. Burgard, S. Thrun “The dynamic window approach to collision avoidance”, in IEEE Robotics & Automation Magazine, vol. 4, no. 1, pp. 23-33 (1997).

[2] M. Dobrevski and D. Skočaj, “Adaptive Dynamic Window Approach for Local Navigation”, 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp.6930-6936 (2020).