

信号機の強化学習における 協調行動の検討

田所 岳之 河原 尊之
東京理科大学工学部電気工学科

1. はじめに

自然界には蟻の餌探索や鳥の群れ飛行など、各個体は単純な行動規則しか持たないが、多数集まると自己組織化された、高度な振る舞いを見せる群知能と呼ばれる技術がある。この考え方を、AI を搭載した機械に持ち込み、AI 同士がローカルに協調することで、全体として新たな能力を会得する試みを「スウォーム AI」と呼称している[1]。本論文ではこの技術を用いて信号機の協調制御について検討した。

2. 実験内容

以下の図 1 のように、16 機の信号、信号機間を基盤の目状に繋ぐ長さ 11 マス・片側 1 マス(両側 2 マス)の道路、2 マス×2 マスの交差点から成る道路でシミュレーションを行った。車は 1 ステップで 1 マス進み、また 2 ステップ毎に 4 台の車が道路端から生成される。車は道路外に出ると消失する。各車は交差点に差し掛かるとそれぞれ 1/3 の確率で直進、左折、右折のいずれかの動作を行う。各車は生成されてから消失するまで、左折、右折をそれぞれ 2 回以上行うことはない。

信号機は強化学習を用いて各方向の交通の時間を決定する。縦交通の青点灯時間のステップ数と横方向の青点灯時間のステップ数を、[縦,横] = [1, 9], [2, 8], [3, 7], [4, 5], [5, 4], [7, 3], [8, 2], [9, 1]のいずれかから選択する。青信号と赤信号の切り替え時には黄色信号が 2 ステップずつ挟まる。上記ステップが終了したら、各信号機は学習を行い、次回行動を決定し実行する。

交通のランダム性のある程度排するため、200 ステップおきに 2 か所の道路を通行止めとした。この条件下で、(a)強化学習なし、(b)強化学習あり・協調なし、(c)強化学習あり・協調ありの 3 パターンの交通を比較した。

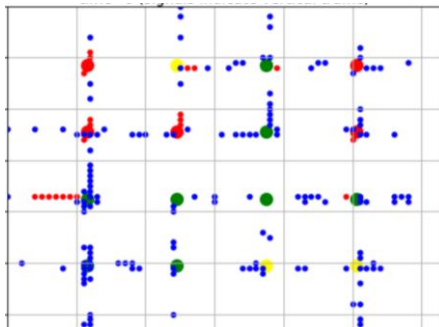


図 1. シミュレーション時の道路環境(小青点は移動車、小赤点は停止車、大点は信号機の縦方向の現示を示す)

3. 各信号機の学習・協調

各信号機には強化学習の手法として Q 学習を用いている。Q 学習とは、ベルマン最適方程式における行動価値関数の方程式を元に編み出された学習方法であり、その更新規則は以下の式で表される。

$$Q^{(i+1)}(s, a) = Q^{(i)}(s, a) + \eta(r + \gamma \max_{a'} Q^{(i)}(s', a') - Q^{(i)}(s, a))$$

ここで s は現在の状態、 s' は次回の状態、 a は現在取った行動、 a' は次回取れる行動、 $Q^{(i+1)}(s, a)$ は更新後の行動価値、 $Q^{(i)}(s, a)$ は更新前の行動価値、 r は今回受け取った報酬である。その他の係数を適切に決定することで、行動価値の収束は早まる。

本実験では各交差点の車の通過数から各交差点で停車した車の総数を引いた数を報酬として与えた。なお、黄色で停止した車や、右折待ちしている車は停止車としてカウントしないこととした。

協調方法として、各信号機は交差点を通り抜けた後すぐの道路を監視し、過去 10 ステップのうち 3 ステップ以上車が通った場合、その車の行き先となる信号機にシグナルを送る。学習時・次回行動決定時にシグナルを受け取っている信号機は、そのシグナルを加味して学習・行動決定を実行することとした。

4. 課題

(a)と比較して(b),(c)は車の総停車数が大きく減少したが、(b)と(c)の間には有意な差が現れなかった。これは、シグナル発信条件が 10 ステップ中 3 ステップ以上の車の通過と条件の幅が広く、同じシグナルが出ている時の中でも実際の交通状態が大きく異なってしまうため、結果としてシグナルを意味のある特徴量として学習にうまく取り込めなかったことが原因であると考えられる。

5. むすび

学習あり・協調なしの時点で大きく停車数が減っているため、現時点では各信号機に AI を搭載すれば、信号機間の協調は不要であるという結論に至った。

参考文献

- [1] 児島丈拓 比江嶋龍也 河原尊之「スウォーム人工知能に基づく複数台ロボットのお協調行動実機検証への環境構築」(2021), ISS-A-065, ISS 特別企画「ジュニア&学生ポスターセッション」