

強化学習を用いた信号現示最適化に関する研究

日本大学 学生会員 ○山崎 颯太 日本大学 正会員 桑原 雅夫
日本大学 正会員 吉岡 慶祐 東北大学 正会員 下川 澄雄

1. はじめに

交通状況に応じて信号制御パラメータを動的に更新する制御方法としては、センサーで得た情報をもとにパラメータ集合を選択する MODERATO 制御などが、従来より運用されている。これに対し、近年では、強化学習を信号制御の最適化に応用しようという研究例^{例えぼ¹⁾}などが見られる。強化学習により信号制御パラメータを決定することで、よりリアルタイムかつ低コストに動的な制御ができる可能性がある。しかし、交通工学的な視点に立ったときの強化学習の適用可能性については、十分な検討が行われているとは言いがたい。例えば、方向別の交通量などの条件が均質でない場合に強化学習により決定されるパラメータが妥当なものか十分な検証が必要である。

そこで、本研究では、強化学習から決定される信号制御パラメータについて、従来の方法による計算結果との比較を通して、交通工学的な実用性を検証することを目的とする。

2. 強化学習の概要

近年、強化学習に関連する研究が様々な分野で進められている。強化学習とは、ある環境の中を自律的に移動するエージェントが、ある状態の時のその行動の結果もたらされる報酬を手がかりに、最適な行動規則を学習する機械学習手法のこと²⁾である。

本研究では、強化学習の一つである Q 学習を用いた。Q 学習を実行するためには、状態、行動、報酬を定義する必要がある。本研究では、離散時間 Δt ごとに状態と行動を定義する。行動は Δt ごとの現示の選択とし、状態は流入路別の待ち台数と前期の現示とする。また、報酬は全流入路の待ち台数の総和と定義した。

$$s = (q_{EW}, q_{WE}, q_{NS}, q_{SN}, a) \quad (1)$$

$$a = \begin{cases} 1 & \text{(EW, WE 方向が青の場合)} \\ 0 & \text{(NS, SN 方向が青の場合)} \end{cases} \quad (2)$$

$$r = -(q_{EW} + q_{WE} + q_{NS} + q_{SN}) \quad (3)$$

ここで、 q_i : i 方向の待ち台数(台)である。

状態 s は式(1)のように、5つの要素を持つベクトルであり、Q 学習で状態 s を記憶するために、 q_i を1台ごとに離散化した。前期の行動 a を状態に含めた理由は、損失時間を考慮するためである。また、エピソード数は 20,000 回から 500,000 回を原則としてサイクル長がある値に収束するまでの回数、計算処理を軽くするため、1step を 5 秒とし、ステップ数は 200、とした。右左折は考慮せず、直進のみを考慮するものとし、全赤時間は 5 秒、損失時間は 5 秒 / 現示である。到着台数の確率変動は考慮せず、一様に到着するものとした。

3. 検討ケース

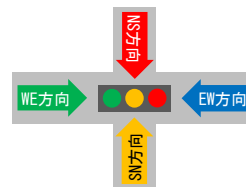


図-1 対象交差点概要

本研究では、図-1に示すような孤立の十字交差点を対象とし、信号現示は単純な 2 現示制御を想定する。各種検討ケースにおける流入交通量および飽和交通流率は表-1に示す通りである。

ケース 1 は基本ケースであり、東西方向・南北方向で交通量・飽和交通流率が等しく、交差点需要率が 0.9 となるケースである。平面交差の計画と設計に示される一様到着の場合の待ち時間を最小化する計算上のサイクル長は 100 秒である。

ケース 2 は、ケース 1 に対して南北方向の交通量が偏在したケースであるが、従来の計算による需要率やサイクル長はケース 1 と同じである。

ケース 3 は、ケース 1 と同じ現示の需要率のまま、SN 方向のみの飽和交通流率・流入台数を変化させたケースである。

ケース 4 は、南北方向の流入台数が時間変動するケースであり、80step まではケース 1 に対して流入台数が小さく、81step 以降はケース 1 と同じ条件としている。

キーワード 強化学習, 信号制御パラメータ

連絡先 〒274-8501 千葉県船橋市習志野台 7-24-1 E-mail : casso19098@g.nihon-u.ac.jp

表-1 流入台数・飽和交通流率と結果

ケース	ケース 1				ケース 2				ケース 3				ケース 4			
方向	EW	WE	NS	SN	EW	WE	NS	SN	EW	WE	NS	SN	EW	WE	NS	SN
流入台数(台/h)	1000	1000	800	800	1000	1000	800	500	1000	1000	800	640	1000	1000	500/800	500/800
飽和交通流率(台/h)	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	1600	2000	2000	2000	2000
現示の需要率	0.5	0.5	0.4	0.4	0.5	0.5	0.4	0.25	0.5	0.5	0.4	0.4	0.5	0.5	0.25/0.4	0.25/0.4
交差点需要率	0.9				0.9				0.9				0.75/0.9			
計算上のサイクル長(秒)	100				100				100				40/100			
計算上の青時間(秒)	50	50	40	40	50	50	40	40	50	50	40	40	20/40	20/40	10/20	10/20
Q学習の青時間(秒)	50	50	40	40	50	50	40	40	50	50	40	40	20/40	20/40	10/20	10/20

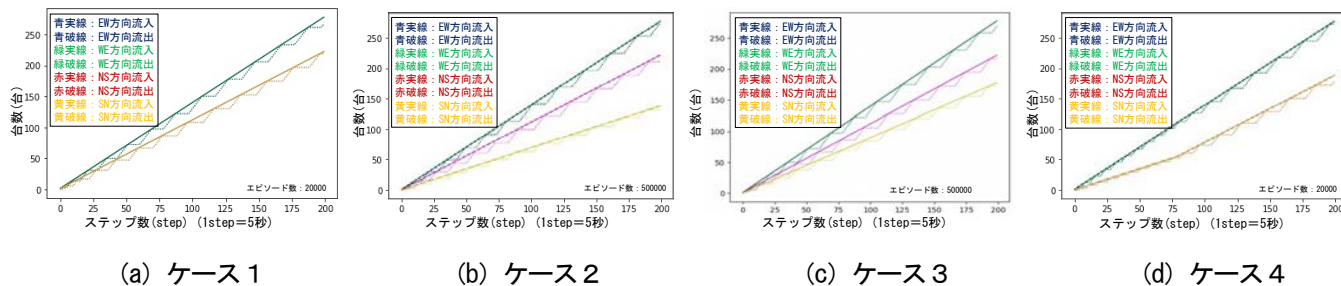


図-2 流入台数と流出台数の関係

4. 検討結果

図-2 (a) ~ (d) は、各ケースにおいて強化学習を実行した結果を、ステップ数と交差点への流入・流出台数の累積値との関係として示したものである。図中の実線は交差点の到着台数、破線は通過台数を示している。また、表-1にこれらの結果をまとめたものを示している。

ケース 1 では、はじめは短い間隔で現示が入れ替わるが、次第に待ち台数が増加し、これに応じてサイクル長も増加している。8 サイクルからは一定のサイクル長 100 秒に収束し、捌け残りなく通過していることがわかる。これは、計算上のサイクル長に等しく、強化学習においても計算上と同じ最適な現示が導出されることが確認された。

次にケース 2 においても、8 サイクル以降は計算上と同じサイクル長 100 秒に収束しており、交通量が偏在しているケースにおいても問題なく計算上と同様の現示を導出可能である。ただし、本ケースにおいては、ケース 1 で最適解が得られたエピソード数 50,000 回の時点では結果が収束せず、待ち台数が増大するようなパラメータとなっており、結果が収束するまでにはエピソード数 500,000 回を要したことも補足する。したがって最適なパラメータを得るには、ケース 1 と比較して学習に時間を要することも確認された。ケース 3 も計算上と同じ現示が得られたが、ケース 2 と同様に収束に必要なエピソード数が増える結果となった。流入台数が変化するケース 4 については、途中の流入台数の

増加に合わせてサイクル長が変化しており、流入台数の時間変化に応じて適切な現示が強化学習により導出されることが確認された。

5. まとめ

本研究では、強化学習により決定された信号制御パラメータと従来の計算結果を比較し、実用上の妥当性を検証した。その結果、到着台数や飽和交通流率の進行方向による偏在や、時間変動の有無に関わらず、従来法と同じ信号制御パラメータを導出したことから、一定の実用性はあることが示された。一方で、到着台数や飽和交通流率の偏在がある場合にはエピソード数を 500,000 回に増やさなければ最適解に収束しないことも明らかとなった。

今後は、車両の到着台数が確率変動する場合にも対応できるかどうかや、複数の交差点を制御する場合にも強化学習を応用できるかなどについて検証を進める予定である。また、Q 学習ではきわめて大きな計算メモリーが必要（次元の呪い）であり、DQN の適用も考えていきたい。

参考文献

- 1) 桑原雅夫, 福田和輝, 橋本申, 佐津川功季, 田名部淳: 強化学習を用いた信号制御パラメータ最適化に関する基礎的研究, 交通工学研究発表会論文集, Vol.42, No.0, pp.563-570, 2022
- 2) 大原剛三: これからの強化学習, 人工知能, Vol.32, No.2, pp.323, 2017