深層強化学習を用いた無信号交差点における協調制御に関する基礎的研究

西嶋　晃佑[1]・片桐　紳太郎[2]・三輪　富生[3]・田代　むつみ[4]・森川　高行[5]

[1]学生会員　名古屋大学大学院工学研究科土木工学専攻　（〒464-8603 名古屋市千種区不老町）
E-mail: nishijima.kosuke.u8@s.mail.nagoya-u.ac.jp

[2]学生会員　名古屋大学大学院工学研究科土木工学専攻　（〒464-8603 名古屋市千種区不老町）
E-mail: katagiri.shintaro.g1@s.mail.nagoya-u.ac.jp

[3]正会員　名古屋大学准教授　未来材料・システム研究所

[4]正会員　名古屋大学大学院　講師　工学研究科土木工学専攻

[5]正会員　名古屋大学教授　未来社会創造機構

本研究では，無信号交差点における協調制御手法を検討した．全ての車両が位置や速度を共有できるコネクティッド自動運転車（CAV）である状況を仮定し，それらがスムーズに合流するための協調制御モデルを深層強化学習により構築し，ミクロシミュレーションにより評価した．無信号 T 字交差点を対象交差点とし，衝突を回避しつつ，旅行時間を最小化できる加速度を遂次的に指示するモデルを，深層強化学習の 1 つである Deep Q-Network と Soft Actor Critic を用いて構築した．分析の結果，本手法では平均速度を高めることができ，効率的な交通制御が可能であることが示された．特に，Soft Actor Critic により構築された協調制御手法は，Deep Q-Network を用いた協調制御より安全で，安定した車両制御が可能であることが示された．また，Soft Actor Critic により構築された協調制御モデルは，優先非優先制御と比較しても，タイムロスが少ない車両制御が可能であることが確認された．

Key Words: *cooperative control, deep reinforcement learning, micro simulation*

1. はじめに

(1) 背景と目的

情報通信技術の発展により，コネクティッドカーの普及に注目が集まっている．国土交通省 [1]によると，2035 年には新車販売の 9 割がコネクティッドカーになると予測されている．コネクティッドカーは，車両の位置や速度をリアルタイムで共有し周囲の交通状況に合わせた協調的な行動を可能とする．また，近年は自動運転技術の進展も著しく，情報通信技術の発展にも伴いコネクティッド自動運転車（CAV, Connected Autonomous Vehicle）が実用化され，これまで実現が不可能であった車両や交通流の制御が可能になると考えられている．特に，5G・低遅延通信の実用化に伴って，コネクティッドカーに関する技術開発が進められており [2,3]，あるエリアを走行する CAV をコントロールサーバーによって制御する時代の到来も予想される．

この CAV の普及を見据えて，交差点の合流調停に関する研究も多く進められている．しかし，これまでの多くの研究では，車両の加減速度を変数とした複雑な数理最適化問題を解くか，協調制御のための車両制御ルールの構築を想定した研究となっている．これらは制御が必要な車両が増えると求解に時間がかかる点や，必ずしも旅行時間を最小化できる解を得るとは限らない点が問題と言える．一方で，機械学習や深層強化学習の進歩は著しく，それらによる交差点の合流調停には大きな可能性があるが，その適応方法や適応効果に関してはまだ知見を蓄積する必要がある．そこで本研究では，深層強化学習による合流調停について，合流調停モデルの構築およびシミュレーション実験を行い，適応可能性に関する基礎情報を得ることを目的としている．

(2) 既往研究と本研究の位置づけ

Dresner and Stone [4]は，交差点内の領域をセルに分割し，各流入車両が通過セルを到着時刻順に予約する制御手法を提案し，流入道路間に交通量の差が少ない状況では，信号制御より効率よく制御できる場合があることを示した．一方，Levin et al. [5]では，車線交通量に偏りがある場

合は効率的な制御ができないことが示されている．Li and Wang [6]は，車両が交差点に進入する順番の組み合わせを網羅的に記した樹形図を基に，車両の交錯が起こらない流入順序の決定による制御手法を提案した．また，Lee and Park [7]は，各流入車両の予定走行軌跡の重複を回避させる車両制御アルゴリズムを提案し，車両が少ない状況下では効率的な制御が可能であることが示されている．Kamal et al. [8]は，車両の衝突リスクの数理最適化問題を解く制御アルゴリズムを提案し，交差点において安全な協調制御ができることを示した．一方，Xu et al. [9]は，交差点制御ゾーンに進入した車両を交錯関係にない車両のみで形成される世代に順番に振り分け，世代順に交差点に進入させる制御手法を提案した．この手法では必ずしも効率的な世代を形成できるとは限らないことに注目し，田代ら [10]は，Xu et al. の制御手法に，制御ゾーンに進入した車両が，最終世代より若い世代に合流可能な場合，車両をその世代に振り分けるアルゴリズムを追加した．また，中垣ら [11]は，制御ゾーンに車両が進入するごとに世代の組み直しを行うアルゴリズムを導入することで，さらに効率化が実現できることを示した．

近年は，機械学習を用いて構築した計算処理が少ない協調制御手法も注目されている．Katagiri et al. [12]は，決定木を用いて交差点へ流入する車両を制御する手法を提案し，深層学習と同程度の精度で制御が可能であることを示している．Isele et al. [13]は，交差点に進入する 1 台の車両を強化学習で構築した学習モデルを用いて制御する手法を提案し，強化学習が協調制御手法に応用できることを示している．また，Yuanyuan et al. [14]は，交差点付近において，各車両を安全な減速度で停止させた際に，他車両と衝突する車両に対して，強化学習で構築した学習モデルを用いて制御する手法を提案し，複数の車両を制御できる学習モデルの構築に成功している．この制御手法は，制限速度が 20km/h 程度である状況では，信号制御よりスムーズな車両制御が可能であることがシミュレーションにより明らかになっているが，走行速度がより高い状況下では，ほぼすべての車両を制御しなければならず，効率的な交通流が実現できない場合が生じる．一方，丹羽 [15]は，常微分方程式で記述され，深層強化学習で最適化された制御モデルを構築し，閉じられた道路ネットワーク内の全車両を衝突を起こさずに制御することを実現した．しかし，実際には道路網は閉じた空間ではないため，対象エリア外との車両の流出入を考慮した制御が可能性について検討が必要といえる．

既往研究を踏まえ，より現実的な協調制御を構築する上での，深層強化学習の利用可能性を検証する．本研究では，対象エリアに流入する全ての CAV 車両を，1 つの中央制御器により制御する状況を想定する．また，独立した 1 つの T 字交差点を対象とする．これは，現実の適用状況を想定すると，丹羽らのように車両を制御する領域を広くとるよりも，交差点単位で制御する方が現実的と考えたためである．制御手法に関しては，Yuanyuan らの手法をベースとするが，より車両速度が高い環境下においても車両制御が可能である手法を深層強化学習により構築し，精度を検証する．

2. 強化学習による車両協調制御の概要

(1) 協調制御の考え方

本研究における協調制御とは，交差点に流入する車両に対して，中央制御器が衝突を回避し，かつ回避行動による遅れが最小となるような加減速指示を与えることを指す．

信号のない T 字交差点に衝突の発生しうる複数の車両が流入する状況を仮定する．この時，すべての車両は CAV で中央制御器と情報的に接続し，交通状況の取得が可能である．図-1 は制御対象車両の決定方法を示している．交差点に接近するすべての車両は中央制御器によって加減速の制御が可能とするが，本研究では，事前の将来予測によって衝突すること判明している車両を制御の対象とする．また，衝突回避のための減速行動によって後続車両と新たに衝突が発生する危険性がある場合，後続車両も次のステップ以降に制御の対象となる可能性がある．ただし，本研究における衝突とは，車両間の車頭距離が 7.5m 以下となった状態と定義する．

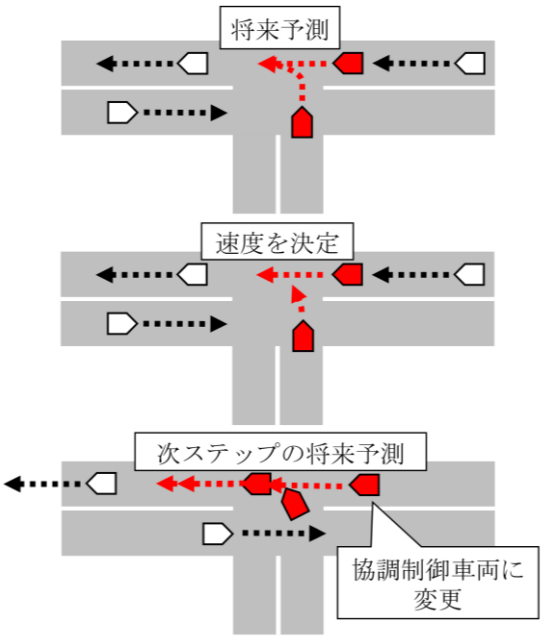


図-1 制御する車両の決定

(2) 深層強化学習

対象車両に最適な加減速指示を行うための制御モデルを，強化学習によって構築する．強化学習とは，行動の主体となるエージェントに一連の行動によって得られる報酬の総和が最大となるような行動パターンを学習させる機械学習手法である．

本研究では，深層強化学習の代表的な手法である Deep Q-Network と Tuomas [16]が提案した Soft Actor Critic を用いて協調制御モデルを構築し，それらの比較を行った．

(3) Deep Q-Network を用いた協調制御

a) 協調制御手法

Deep Q-Network ではエージェントが状況に応じて最も価値が高い行動をとることが出来るように，ある状態sにおけるエージェントの指令の価値を推定可能なニューラルネットワークを構築する．

図-2 は協調制御に用いる Deep Q-Network の構造を示している．本研究におけるエージェントとは中央制御システムを指し，指令とは，各車両に与える加減速度の組み合わせを示す．つまり，協調制御において，中央制御システムはあらかじめ設定された車両の加減速指示の候補の中から状況に応じた最適な指示を行う．例えば，行動価値が表-1 となる状態では，25 パターンある加減速度の組み合わせ候補の中から，車両 1 には$0.0\,m/s^2$，車両 2 に$1.0\,m/s^2$の加減速度の組み合わせを指示する．加速度の組み合わせ総数は，車両の台数nと各車両がとり得る離散的な加速度の選択肢数mを用いて，m^nで表される．ただし，本研究では車両に指示可能な加速度を-1.0, 0.5, 0.0, -0.5, -1.0 m/s^2の 5 パターンとする(m=5)．

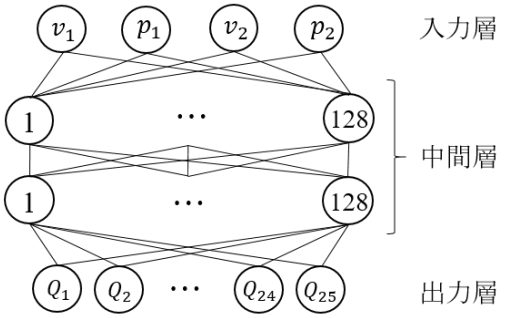


図-2 Deep Q-Network でのニューラルネットワーク

表-1 ある状態での行動価値

		車両 1				
	加減速度 (m/s^2)	1.0	0.5	0.0	-0.5	-1.0
車両 2	1.0	-0.4	0.1	1.7	0.6	0.2
	0.5	1.4	1.1	0.5	-0.1	0.1
	0.0	-0.7	-0.4	-0.2	0.0	0.9
	-0.5	1.1	0.7	0.3	0.2	0.4
	-1.0	0.5	0.4	0.4	0.1	-0.1

b) 学習の流れと行動価値Qの更新

交差点に車両が流入する前の状態を初期状態とし，ランダムに車両を流入させる．すべて車両が交差点を通過するか，衝突の発生，制限時間を超過した場合を終了条件として，一回のシミュレーションとする．学習では，モデルによる行動選択を行いながらシミュレーションを行い，実際に行動によって得られた報酬とモデルによる推定値の誤差が小さくなるように，モデルのパラメータを更新する．モデルは状態変数として全ての車両の位置(p)と速度(v)を入力値とし，それぞれ 128 ノードで構成される 2 層の中間層をもつ．

また，同時に学習シミュレーションでは，累計報酬の総和が大きくなるような行動の探索を行い，行動価値を更新する．式(1)の様に，時刻tにおける状態s_tでの指令A_tの行動価値$Q(s_t, A_t)$の更新では，学習率αを設定し，従来の行動価値と，行動により得た即時報酬rと将来受け取る報酬の合算値とによって更新する．本研究では，即時報酬rを表-2 のように設定する．ステップコストに対する即時報酬が負値であるため，衝突を回避できても旅行時間が長くなる行動をとると，最終的な報酬が減少する．また，衝突の発生や制限時間を超えるような長時間の旅行時間でも報酬が大きく減少する．このようにして衝突が起こらず，交差点内車両の平均速度が大きくなるよう学習を行う．

$$Q(s_t, A_t) = (1-\alpha)Q(s_t, A_t) + \alpha\left(r + \gamma \max_{a_{t+1}} Q(s_{t+1}, A_{t+1})\right) \quad (1)$$

ここに，αは学習率，rは行動により得た即時報酬，γは割引率である．割引率γは，時間の経過に伴い，将来受け取る報酬を指数的に減衰させるための数値である．

学習ではモデルによって推定された行動価値が高い指令を取り続けると探索が不十分なまま，局所解に収束する場合がある．そのため ε-greedy 探索法を採用し [17]，ランダム行動確率εを用いて，モデルによる推定結果を用いずランダムに行動させ，探索範囲を拡大させた．また，一回のシミュレーションにより得られたデータのみを用いて学習すると，その状況に過敏に反応し学習が安定しない場合がある．経験再生(Experimental Replay)によって蓄積した過去のシミュレーションデータを含めたデータセットからランダムにデータを抽出して学習することで [18]，学習の安定化を図る．

表-2 Deep Q-Network による協調制御の即時報酬

ステップコスト	衝突	全車両が交差点を通過	制限時間経過
-1	-400	400	-400

c) 学習設定

Deep Q-Network では，T 字一車線交差点に進入する車両数を 3 台の場合と 4 台の場合を設定した．学習に関するパラメータは表-3 の通りである．

学習率αと割引率γは，式(1)を用いて行動価値を更新するために使用される数値であり，バッチサイズとは，バッファからランダムで取り出すデータセットの値である．エピソード数について，1 エピソードとは全車両が対象とする道路ネットワークに進入してから通過するか，総旅行時間が制限時間に達するまでのことであり，制御モデルを学習させるために実行したエピソードの回数である．総旅行時間の制限時間は，3 台制御では 200 秒，4 台制御では 300 秒とした．また，イプシロン値εは，初期値を 1 として，3 台制御の場合は 200 エピソード毎，4 台制御の場合は 500 エピソード毎に値を 0.99 倍した．

表-3　Deep Q-Network の学習パラメータ

	学習率 α	割引率 γ	バッチサイズ (セット)	エピソード数 (回)
3 台制御	0.0005	0.98	32	100,000
4 台制御				250,000

(4) Soft Actor Critic を用いた協調制御

a) 協調制御手法

Soft Actor Critic では，エージェントが各車両に対して状況に応じて最も価値が高くなりやすい行動を指示出来るように，方策を決定する Actor と方策を評価する Critic から成るニューラルネットワークを構築する．図-3 は，車両台数n台でのニューラルネットワークを表している．Actor では，ある状態sにおける各車両の行動確率を決定する（図-3 左上）．この行動確率は，全ての車両ごとに与えられ，選択する加減速度の確率分布で表現できる．本研究では，行動確率を正規分布$N(\mu, \sigma^2)$に従う確率分布で表現する．Critic では，ある状態sで行動確率を基にして選択された加減速度aの行動価値を推定する（図-3 右下）．Soft Actor Critic では，行動価値は指定した加減速度の範囲内において連続値で表現され，方策を評価する際に必要とされる．

Soft Actor Critic では，方策πを用いて各車両に加減速度aを指示する．方策πには，あらゆる状態での行動確率が集約されており，エージェントは各車両に対して -1.0 ~ 1.0m/s^2 の中から行動確率が最も高い加減速度を指示する．特に，Deep Q-Network を用いた協調制御と異なり，加減速度を連続値で指示できる点が Soft Actor Critic の特徴である．

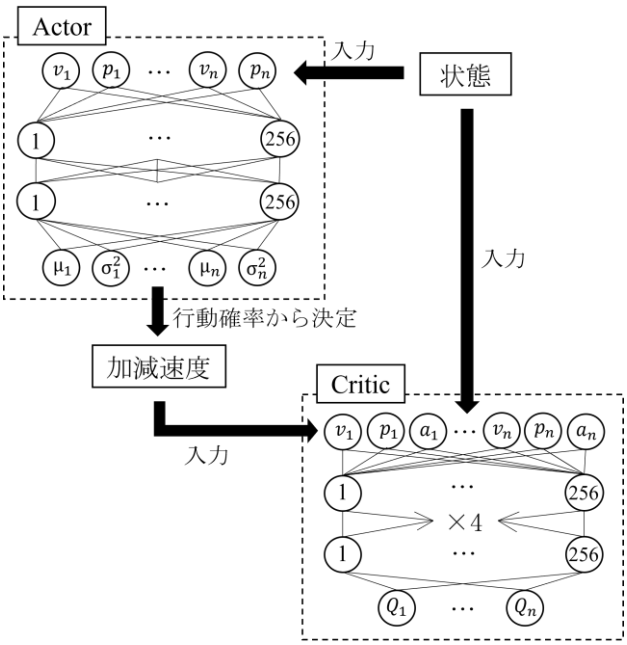


図-3　Soft Actor Critic でのニューラルネットワーク

b) 協調制御の学習

Soft Actor Critic では，Deep Q-Network と同様のシミュレーション環境下で，各車両のあらゆる状態での行動価値が最大化されるような，方策πを探索することで制御モデルを構築する．つまり，エージェントが各車両に対して，行動価値が最大化されるような加減速度を指示できるように，行動確率の変数μ, σ^2を最適化させる．

学習における初期設定では，各車両の方策は正規分布$N(0, \infty)$に従うとする．その後エージェントは，協調制御を指示された車両に対して，暫定的な方策に従って加減速度を指示し，繰り返し車両制御を実行する．車両を協調制御した際の，エージェントが車両に指示した加減速度，行動により受け取った即時報酬，行動前後の全車両の位置・速度は，1 つのデータセットとしてバッファに保管する．その後，1 エピソードごとに幾つかのデータセットをバッファから取り出し，そのデータセットを用いて，各車両に指示した加減速度の行動価値を更新する．Soft Actor Critic を用いた協調制御では，時刻tにおける状態s_tでの加減速度a_tの行動価値$Q(s_t, a_t)$は，式(2)のように更新する．

$$Q(s_t, a_t) = r + \gamma E_\pi[Q(s_{t+1}, a_{t+1})] + \alpha E_\pi[-\log \pi(a_{t+1}|s_{t+1})] \quad (2)$$

ここに，rは行動により得た即時報酬，γは割引率，αはエントロピー項の反映率である．即時報酬は，Deep Q-Network と同様に表-2 のように設定し，割引率γを 0.99 に設定した．$\pi(a|s)$は，方策に従ってエージェントがある状態sにおいて方策aを指示する確率である．式(2)の右辺第二項は，暫定の方策に従った際に今後得られる報

酬の期待値を表し，右辺第三項は，方策エントロピー項を表している．方策エントロピーとは，加減速度の対数行動選択確率$\log \pi(a|s)$の期待値であり，行動選択確率の分散が大きくなにつれてエントロピー項も大きくなる．エントロピー項の役割としては，方策の分散を大きくすることで，様々な加減速度選択させ，エージェントに新しい最適解を探索させることにある．

Soft Actor Critic を用いた協調制御では，方策に従った際に得られる行動価値を最大化させることで制御を構築した．本研究では，それぞれの状態での，方策により決定される行動価値を最大化させるために，式(3)を用いて方策を繰り返し更新させる．

$$\pi_{new} = \arg\min_{\pi' \in \Pi} D_{KL}\left(\pi'(.|s_t)||\frac{\exp\left(\frac{1}{\alpha}Q^{\pi_{old}}(s_t,.)\right)}{Z^{\pi_{old}}(s_t)}\right) \quad (3)$$

ここに，αはエントロピー項の反映率，D_{KL}はKLダイバージェンス，$Z^{\pi_{old}}(s_t)$は状態s_t，方策π_{old}での分割関数である．$Q^{\pi_{old}}(s_t,.)$は，状態s_tにおいて，暫定の方策π_{old}に従って加減速度を指示した際に受け取る行動価値を表しており，$\pi'(.|s_t)$は，状態s_tにおける，Πに属す任意の方策を表す．Tuomas et al. は，式 (3) を用いて方策を最適化させると方策に従った際に得られる行動価値が最大化できることを証明している．

c) 学習設定

Soft Actor Critic では，T字一車線交差点に進入する 4 台の協調制御手法と T 字二車線交差点に進入する 4 台の協調制御手法を構築した．学習に関するパラメータは表-4 の通りに設定した．

エントロピー項の反映率αと割引率γは，式(2)を用いて行動価値を更新するために使用される数値である．二車線交差点の制御学習で用いたエントロピー項の反映率αは，Bayesian 最適化によって推定された最適解を適用した [19]．エピソード回数については指定せず，制御モデルが構築されたと判断されたタイミングで学習を終了させた．この際，制御学習を行う 100 エピソードごとに衝突率を出力させ，値が 0%であれば学習を中断し，5000 回の評価シミュレーションを行う．この評価シミュレーションにおいて一度も衝突が起こらなければ制御モデルが構築できたと判断し，学習を終了した．もし，5000 回のシミュレーションで一度でも衝突が起これば中断した学習を再開させた．また，総旅行時間の制限時間を 300 秒に設定した．

表-4　Soft Actor Critic での T 字交差点制御の学習パラメータ

	エントロピー項の反映率α	割引率γ	バッチサイズ（セット）
一車線交差点	0.400	0.99	256
二車線交差点	0.619		

3. 学習シミュレーション設定

本研究では，学習フェーズで協調制御手法を構築し，シミュレーションフェーズで構築した手法を評価した．学習環境とシミュレーション環境は同一とし， 道路ネットワークは 2 種類を用いることとする．一つ目の道路ネットワークは，図-4 の通り，一方通行の道路に他の一方通行道路が接続する T 字交差点で，一辺の長さを 150 m とした．横方向の道路を制限速度 15 m/s（54 km/h）の幹線道路とし，縦方向の道路を制限速度 10m/s（36km/h）の生活道路とした．また、生活道路からの流入に際して一旦停止は行わないが，生活道路から交差点を左折する際は，車両の速度が 5 m/s（18km/h）以下になるよう減速を行う．また，道路ネットワークに進入する際の車両の速度は，幹線道路から進入する車両は 10m/s，生活道路進入する車両は 7m/s とした．

2 つ目の道路ネットワークも同様に，T 字交差点であるが，図-5 の通り，一辺の長さを 200m の両側二車線道路で構成されている．全ての車線の制限速度を 15m/s とし，右左折時の速度が 5m/s 以下になるように減速を行う．減速の方法は，T 字一車線交差点の道路ネットワークと同様である．また，道路ネットワークに進入する際の車両の速度は，一律で 10m/s とした．なお，交差点内部での衝突判定においては，交差点内にある車両の軌跡を調査し，図-6 の走行軌跡の交錯が発生しない車両同士は衝突判定しない．

いずれの環境でも，交差点内の車両は，車両同士の衝突が発生するランダムなタイミングでネットワークに進入する．

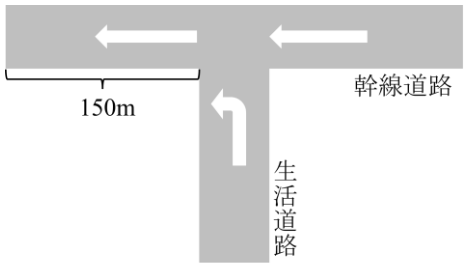


図-4　T 字一車線交差点の道路ネットワーク

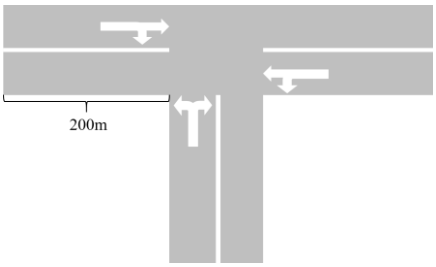


図-5　T 字二車線交差点の道路ネットワーク

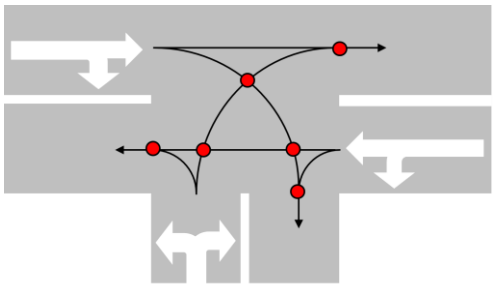


図-6　車両の交錯関係

4.　シミュレーション結果

構築した制御モデルはシミュレーション実験により評価する．車両の流入時刻をランダムに変化させつつ，シ 5000 回のシミュレーションによるモデル性能の検証を行った．

(1)　T 字一車線交差点での 3 台の協調制御

Deep Q-Network を用いて，車両が幹線道路から 2 台，生活道路から 1 台交差点に進入する際の協調制御手法を構築した．制御学習時の 100 エピソード毎の衝突率は図-7，シミュレーション結果は表-5 の通りとなった．

図からも分かるように，学習を重ねるにつれて衝突率が減少しており，制御モデルが車両間の衝突を回避するような制御方法を学習している様子が示されている．

シミュレーション評価を行った結果，衝突率は 0%となり安全な制御手法が構築できたといえる．また，平均速度は，9.83km/h となり，シミュレーションのエピソード別の総旅行時間分布は図-8 のようになった．多くは，

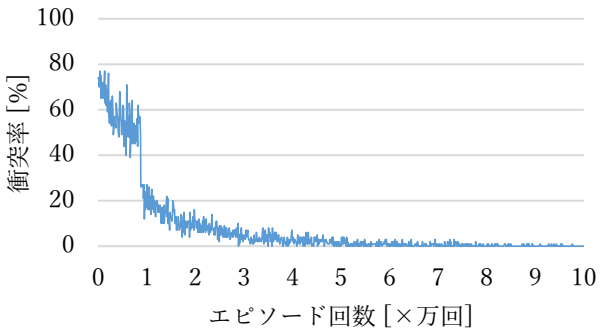


図-7　Deep Q-Network での 3 台制御学習時の 100 エピソード毎の衝突率

表-5　Deep Q-Network での 3 台制御手法のシミュレーション結果

	衝突率 (%)	平均速度 (m/s)	学習時間 (h)
シミュレーション結果	0.00	9.83	6

※計算環境：Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz，メモリ 8GB

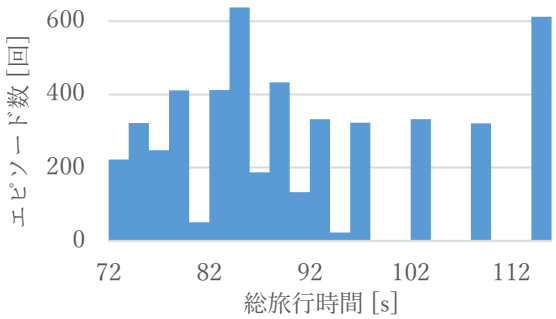


図-8　Deep Q-Network での 3 台制御シミュレーションの総旅行時間分布

総旅行時間が 72 秒から 98 秒程度に収まっているが，114 秒から 116 秒の総旅行時間となるエピソードが 600 回以上存在している．このことから，衝突は発生しないものの 一部の車両流入パターンでは過剰な減速制御等による遅れが発生しているものと考えられる．

(2)　T 字一車線交差点での 4 台の協調制御

Deep Q-Network と Soft Actor Critic の性能を調査するため，幹線道路から 2 台の車両，生活道路から 2 台の車両が交差点に進入する状況に対してそれぞれ学習を行った．Deep Q-Network を用いた協調制御手法の制御学習時における 500 エピソード毎の衝突率を図-9 に，Soft Actor Critic を用いた協調制御手法の制御学習時における 100 エピソード毎の衝突率を図-10 に示す．また，それぞれの制御手法と優先・非優先制御でのシミュレーション結果を表-6 に示す．

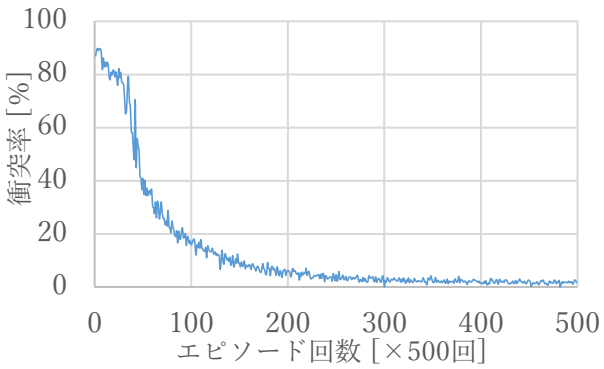


図-9　Deep Q-Network での 4 台制御学習時の
500 エピソード毎の衝突率

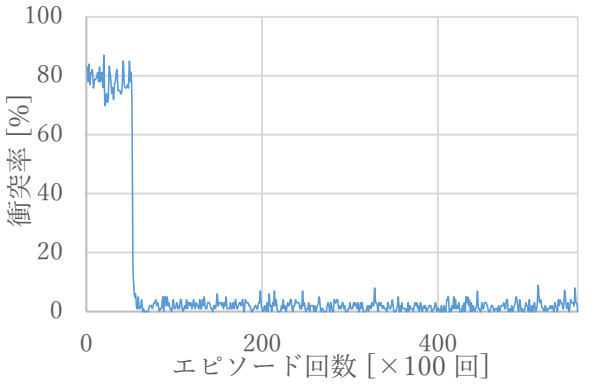


図-10　Soft Actor Critic での 4 台制御学習時の
100 エピソード毎の衝突率

Deep Q-Network を用いた協調制御手法では，衝突率は緩やかに低下する傾向にあり，制御モデルを構築するまでに多くのエピソードが必要であった．これは，制御モデルの学習を安定させるために，緩やかにエージェントがランダムに行動を選択する確率を低下させたことに起因する．Soft Actor Critic を用いた協調制御手法では，エピソード回数 5000 回付近で衝突率が急激に低下したが，その後学習が終了するまでに多くのシミュレーションを要した．また，Deep Q-Network を用いた協調制御手法は，Soft Actor Critic と比較して，制御の構築に要した学習エピソード回数が 4 倍以上多くなった．このように，学習手法の違いにより学習時のプロセスや効率性が異なることが分かる．

表-6 よりシミュレーション結果について比較すると，Soft Actor Critic を用いた協調制御手法は，平均速度が Deep Q-Network を用いた協調制御手法と優先・非優先制御より高く，また衝突率が 0%であるため，精度の高い協調制御が実現できていることが分かる．平均速度の分散は，Soft Actor Critic の制御モデルが，Deep Q-Network の制御モデルより小さな値となった．そのため，Soft Actor Critic を用いたモデルは，様々な状況で安定的な速度で

表-6　4 台制御手法のシミュレーション結果

	衝突率（%）	平均速度（m/s）	平均速度の分散	学習時間（h）
Deep Q-Network	1.08	9.30	0.70	14
Soft Actor Critic	0.00	11.50	0.15	5
優先・非優先制御	0.00	9.18	0.15	

※計算環境：Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz，メモリ 8GB

車両制御を行えることが分かる．また，Deep Q-Network を用いた協調制御手法では，衝突率を 0%にすることが出来なかった．Deep Q-Network による協調制御を確認するため，図-11 に Deep Q-Network を用いた制御手法のシミュレーションにおいて衝突が発生したエピソードの車両挙動を示す．

衝突は，エージェントが車両 2 と車両 3 の衝突回避方法について十分に学習が行えなかったことにより発生したと考える．図-11 の通り，衝突を初めて検知した際に，車両 2 を減速させることで衝突回避を試みたが，これにより車両 2 と車両 3 が接近し，車両の衝突が発生した．Deep Q-Network を用いた協調制御手法の様に，離散的な加速度しか指示できない制御モデルでは，加速度の最適解を指示できないため，このような 3 台以上の車両が関連する衝突を回避することは困難であると考える．

続いて，Deep Q-Network と Soft Actor Critic を用いた協調制御手法のシミュレーションでの総旅行時間分布をそれぞれ図-12，図-13 に示す．Deep Q-Network の制御モデルでは，3 台制御の場合（図-8）と同様にばらつきが非常に大きいことがわかる．そのため，この制御モデルは，様々な状況に対応できる時間効率の良い車両制御が出来ていないと考えられる．この要因としては，Deep Q-Network の制御モデルではエージェントが車両に対して離散的な加減速度でしか指示できない点が挙げられる．また，制御する車両台数が増加すると，ニューラルネットワークの出力層が指数的に増加するため，車両台数が多い環境下では，Deep Q-Network の制御モデルの学習が複雑化される点も考慮されたい．

一方，Soft Actor Critic を用いた制御モデルの総旅行時間分布は 100~101 秒の範囲が最も多く，総旅行時間が大きくなるにつれ分布が低下しており，Deep Q-Network を用いた制御モデルと比較して，ばらつきがかなり小さいことが分かる．これより，Soft Actor Critic を用いた制御モデルは総旅行時間のロスが起こりにくく，効率的な制御が可能となるといえる．

(3) T 字二車線交差点での 4 台の協調制御

以上を踏まえ，T 字二車線交差点において，横方向道路の左から 2 台，右から 1 台，縦方向道路の下から 1 台が交差点に進入する状況に対して，Soft Actor Critic を用いて協調制御を学習した．本研究では，交差点での車両の進行方向が固定した環境で制御の学習とシミュレーションを実行した．図-14 の様に，横方向道路の左右から進入する車両は直進し，縦方向道路の下から進入する車両は右折する．また，今までの車両の許容加減速度は$\pm 1.0 m/s^2$としたが，本研究の制御では，$\pm 2.0 m/s^2$とした．シミュレーションの結果は，表-7 の通りになった．

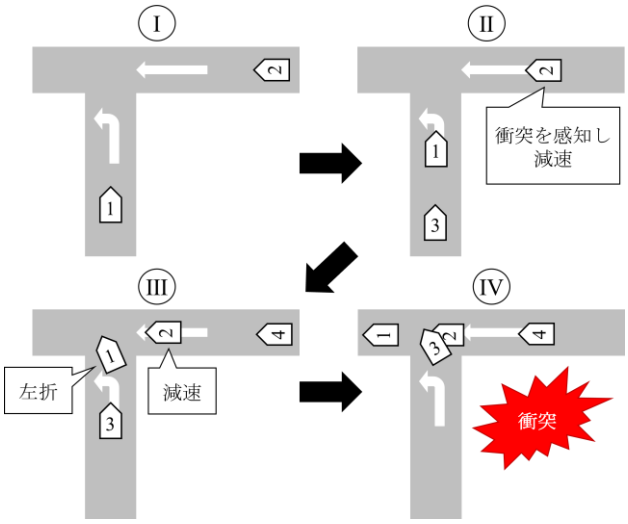


図-11　Deep Q-Network を用いた 4 台制御での衝突事例

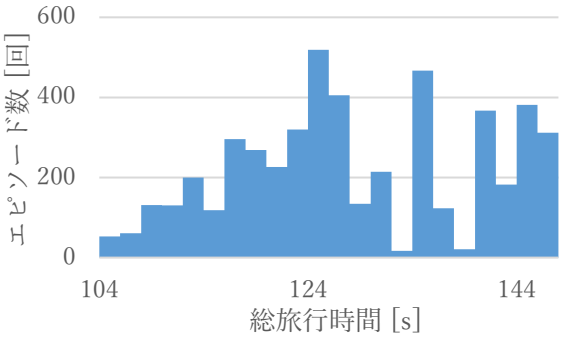


図-12　Deep Q-Network 制御モデルの
4 台制御シミュレーションの総旅行時間分布

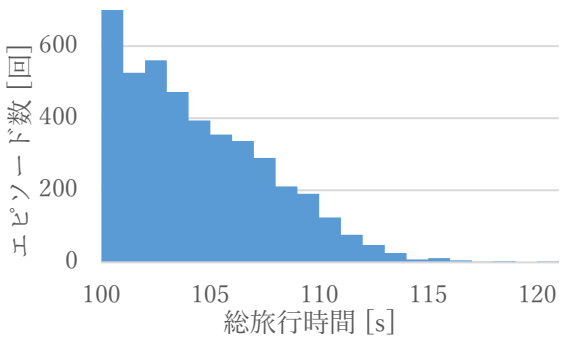


図-13　Soft Actor Critic 制御モデルの
4 台制御シミュレーションの総旅行時間分布

同数の車両が道路ネットワークに流入する場合，車線数の多い T 字二車線道路は，T 字一車線道路と比較して車両同士が衝突するリスクが低くなる．そのため，衝突率は制御モデル学習前から既に 15%程度であった．制御モデルは，2300 エピソード学習することにより，15%程度ある衝突率をさらに 0%に抑えた車両制御を行えることが分かった．

総旅行時間分布は，図-15 となった． 総旅行時間はその値が大きくなるにつれ頻度が低下しており，Soft Actor Critic により構築した T 字一車線交差点の協調制御モデルと同様に，総旅行時間分布のばらつきが小さいことが分かる．これより，Soft Actor Critic により構築した協調制御モデルは，T 字二車線交差点において，効率的でかつ安定した車両制御が可能であることが分かった．

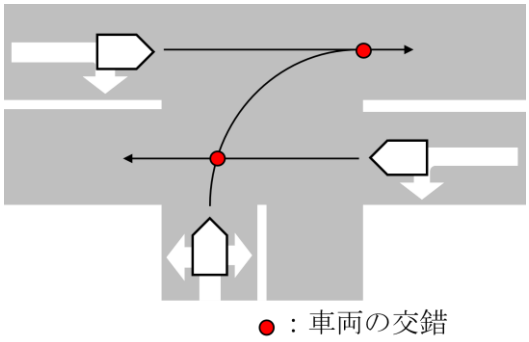


図-14　交差点内の車両の走行軌道

表-7　二車線交差点でのシミュレーション結果

衝突率 (%)	平均速度 (m/s)	学習エピソード数（回）	学習時間 (h)
0.00	14.41	2300	1

※計算環境：Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz，メモリ 8GB

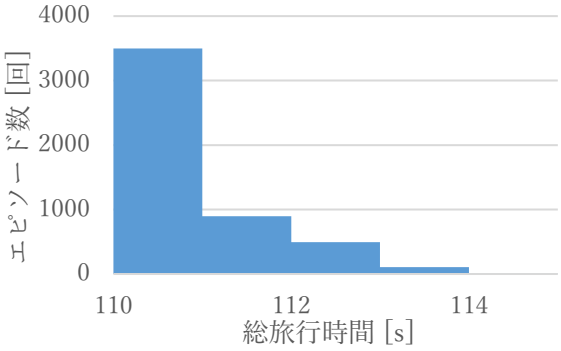


図-15　Soft Actor Critic 二車線制御での総旅行時間分布

5. 結論

(1) 本研究で得られた知見

本研究では，Deep Q-Network と Soft Actor Critic を用いて，無信号T字交差点における協調制御モデルを構築し，ミクロシミュレーションによりそれぞれのモデルを比較，検討した．深層強化学習により構築した協調制御モデルは，徐行以上の速度で走行する車両のT字一車線交差点における安全な車両制御が可能であることがシミュレーションにより分かった．

Deep Q-Network を用いた制御モデルは，T 字一車線交差点付近に車両が 3 台存在している場合は，衝突率が 0%の車両制御が可能であるが，4 台以上存在している場合は安全な車両制御が行えなかった．一方で，Soft Actor Critic により構築した制御モデルを用いれば，T 字一車線交差点付近に車両が4台存在しても，効率的かつ衝突率が 0%の車両制御が可能であることが分かった．また，T字二車線交差点付近に交差点での進行方向が固定された車両が4台存在しても，効率的かつ衝突率が 0%の車両制御が可能であると示された．

(2) 今後の課題

本研究では，主にT字一車線交差点での協調制御を行ったが，今後は，T字二車線交差点の効率的な制御モデルを構築することを考えている．本研究でも，T字二車線交差点における制御モデルを構築したが，車両台数が少ない環境のみしか制御が行えない．今後は，同時に制御できる車両の数を増やし，車両が右左折様々な挙動をとる環境において効率的な車両制御が行える制御モデルの構築を目指す．

また，本研究では，学習時でもシミュレーション時でも，指定した車両が道路ネットワークに進入してから通過するまでを 1 エピソードとし，エピソードごとに分離して制御モデルを動かしている．この状況で制御モデルを構築しても，実際の交差点の様に連続的に車両が進入する状況では効率的な車両制御が出来ない可能性がある．そのため，今後は，道路ネットワークに連続して車両が進入する状況で制御を学習させ構築すべきである．

謝辞：本研究は，JSPS 科研費 19H2260 の助成を受けたものです．

REFERENCES

1) 国土交通省：主要データ集，第 73 回基本政策部会配布資料, pp.8-12, 2020. [Ministry of Land, Infrastructure, Transport and Tourism: *Major data collection, The 73rd meeting of the Ministry of Fundamental Policies to distribute cloth information*, pp.8-12, 2020.]
2) トヨタ自動車：トヨタ自動車、ITS 専用周波数を活用した世界初の運転支援システム ITS Connect を新型車に採用，トヨタ自動車 HP，2015. [Toyota Motor Corporation: *Toyota Motor Corporation has adopted ITS Connect, the world's first driver assistance system that utilizes ITS-dedicated frequencies, in its new models*, Toyota Motor Corporation HP, 2015.]
3) 吉野仁：５Ｇ高信頼・低遅延通信を用いたコネクティッド・カーおよび自動運転実現に向けた実証実験 -スマート交差点・遠隔運転への５Ｇの応用-，第 18 回 ITS シンポジウム，2020. [Yoshino, J.: *Demonstration Experiment for Realization of Connected Car and Automatic Driving Using 5G Highly Reliable and Low Latency Communications - Application of 5G to Smart Intersection and Remote Driving-*, *The 18th ITS Symposium*, 2020.]
4) Dresner, K. and Stone, P.: Multiagent traffic management: a reservation-based intersection control mechanism, Proc. AAMAS, New York, NY, USA, pp.530-537, 2004.
5) Levin, M.W., Boyles, S.D. and Patel, R.: Paradoxes of reservation-based intersection controls in traffic networks, Transportation Research Part A: Policy and Practice, Volume 90, August 2016, pp. 14-25, 2016.
6) Li, L. and Wang, F.Y.: Cooperative driving at blind crossings using intervehicle communication, IEEE Transaction on Vehicular Technology, Vol. 55, No.6, pp.1712-1724, 2006.
7) Lee, J. and Park, B.: Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment, IEEE Transaction on Intelligent Transportation, Vol. 13, No.1, pp.81-90，2012.
8) Kamal, M.A.S., Imura, J, Hayakawa, T, Ohata, A, Hayakawa, T. and Aihara, K.: A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights, IEEE Transaction on Intelligent Transportation, Vol. 16, No.3, pp.1136-1147, 2015.
9) Xu, B., Li, E.S., Bian, Y., Lib, S., Ban, X.J., Wang, J. and Li, K.: Distributed conflict-free cooperation for multiple connected vehicles at unsignalized intersections, Transportation Research Part C: Emerging Technologies, Vol 93, pp.322-334, 2018.
10) 田代むつみ，施展華，三輪冨生，森川高行：コネクティッドカーの交差点協調制御に関する研究，土木計画学研究・講演集，Vol.61，CD-ROM，2021. [Tashiro, M., Shi, Z., Miwa, T. and Morikawa, T.: *Research on Cooperative Control at an intersection for Connected Cars*, *Civil Engineering Planning Research and Proceedings*, Vol.61, CD-ROM, 2021.]
11) 中垣弦一郎，田代むつみ，三輪冨生，森川高行：右折レーンを活用した無信号交差点における協調制御，第 42 回交通工学研究発表会，2022/8/9 - 2022/8/10. [Nakagaki, G., Tashiro, M., Miwa, T. and Morikawa, T.: cooperative Control Utilizing Right Turn Lane at non-signalized Intersection, *The 42nd Conference on Japan Society of Traffic Engineers*, 2022/8/9 - 2022/8/10.]
12) Katagiri, S., Miwa, T., Tashiro, M. and Morikawa, T.: Effect of optimization time-scale on learning-based cooperative merging control at a non-signalized intersection, IEEE Access, Vol. 11, pp. 32857-32868, 2023. (DOI: 10.1109/ACCESS.2023.3263118)

13) Isele, D., Rahimi, R., Cosgun, A., Subramanian, K. and Fujimura, K.: Navigating Occluded Intersections with Autonomous Vehicles using Deep Reinforcement Learning , IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 2034-2039, 2018.
14) Yuanyuan, W., Haipeng, C. and Feng, Z.: Decentralized coordination learning of autonomous intersection management for connected and automated vehicles, ELSEVIER Transportation Research Part C 103, pp.246-260, 2019.
15) 丹羽健太：デジタルツインでモビリティ群を賢く制御する-分散深層学習がもたらす未来の可能性，NTT 技術ジャーナル，Vol.34，No.8，pp.45-49，2022. [Niwa, K.: *Digital Twin for Intelligent Control of Mobility Swarms-The Future Possibilities of Distributed Deep Learning*, *NTT Technology Journal*, vol.34, No.8, pp.45-49, 2022.]
16) Tuomas, H., Aurick, Z., Pieter, A. and Sergey, L.: Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, arXiv:1801.01290, 2018.
17) 斎藤康毅：ゼロから作る Deep Learning ④ - 強化学習編，株式会社オライリー・ジャパン，2022. [Saitoh, K.: *Deep Learning from scratch (4) - Reinforcement Learning*, O'Reilly Japan, Inc., 2022.]
18) 久保隆宏：Python で学ぶ強化学習 [改訂第 2 版] 入門から実践まで，株式会社講談社，2019. [Kubo, T.: *Learning Reinforcement Learning using Python [Revised 2nd Edition] From Introduction to Practice*, KODANSHA Ltd., 2019.]
19) T. サージョス（石川達也ほか訳）：機械学習 ベイズと最適化の観点から[原著第２版]，共立出版株式会社, 2022. [Sergios, T. (translated by Ishikawa, T.): *Machine Learning a Bayesian and Optimization Perspective 2nd ed.*, Kyoritsu Shuppan Co., Ltd., 2022.]
20) Tuomas, H., Aurick, Z., Kristian, H., George, T., Sehoon, H., Jie, T., Vikash, K., Henry, Z., Abhishek, G., Pieter, A. and Sergey, L.: Soft Actor-Critic Algorithms and Applications, arXiv:1812.05905, 2018.

(Received ?, 2023)
(Accepted ?, 2023)

A PRELIMINARY STUDY ON COOPERATIVE CONTROL USING DEEP REINFORCEMENT LEARNING AT NONSIGNALIZED INTERSECTION

Kosuke NISHIJIMA, Shintaro KATAGIRI, Tomio MIWA, Mutsumi TASHIRO and Takayuki MORIKAWA

In this study, a cooperative control method at a nonsignalized intersection are examined. Assuming that all vehicles are connected cars, cooperative control models for their smooth merging is constructed using deep reinforcement learning and evaluated using micro-simulation. Using Deep Q-Network and Soft Actor Critic, a type of deep reinforcement learning, a cooperative control model at a nonsignalized T-intersection that minimizes travel time while avoiding collision was constructed. The analysis results showed that this method is capable of efficient traffic control with high average speed. In particular, it was shown that the cooperative control method constructed by Soft Actor Critic can control traffic more safely and stably than Deep Q-Network one. In addition, it was confirmed that the cooperative control model constructed by Soft Actor Critic can control vehicles without collision and with less time loss than the priority non-priority control.