強化学習による交差点混雑に着目した協調的経路計画システムの基礎的検討

片桐　紳太郎[1]・三輪　富生[2]・森川　高行[3]

[1]学生会員　名古屋大学大学院　工学研究科土木工学専攻（〒448-8603 名古屋市千種区不老町）
E-mail: katagiri.shintaro@s.mail.nagoya-u.ac.jp

[2]正会員　名古屋大学准教授　未来材料・システム研究所
[3]正会員　名古屋大学教授　未来社会創造機構

従来の経路配分計画では、交通流を巨視的にとらえて分析を行うことがほとんどあり，交差点内での走行位置の重複のような微視的事象を考慮していない．そこで本研究では，交差点交通流における経路交錯による遅れに着目し，交差点内での微視的現象を考慮した経路計画手法についての検討を行った．メソスコピック交通流シミュレータを構築し，交差点に仮想的に設置された経路計画システムが，流入する車両に経路を割り当てることができる環境を想定し，強化学習を用いて動的経路計画モデルを構築した．実験の結果，学習した経路計画モデルは，車両の目的地に応じて他車両との走行軌跡の交錯が減るような経路計画を行うことが可能であり，対象交通流の円滑性が向上することが示された．

Key Words: *route allocation,connected and autonomous vehicles,reinforcement learning,conflict*

1. はじめに

近年，Conneceted and Autonomous Vehicles(CAVs)による車両間の協力的行動を利用した交通流の管理・計画手法に注目が集まっている．中でも，交通流における遅れや渋滞の発生の主要因となる交差点交通流を CAVs や交通インフラからリアルタイムに共有される交通情報をもとに車両を制御し，安全かつ効率的な合流手法に関する研究が盛んにおこなわれている[1)-8)]．

Dresner and Stone [1)]は車両の衝突を回避するため，各車両が交差点内での走行位置を予約する手法を提案した．また，Xu et al [2)]は同時に交差点に進入する車両群を世代として定義し，同一世代内で交差点内での走行軌跡の交錯が発生しないような世代の構成方法を示した．また，多くの研究で車両の位置や速度を変数として，交通流を数理モデルとして表現し，衝突を回避しかつ総旅行時間を最小化するような車両の加速度制御の最適解を，最適化問題の定式化や強化学習によって求める手法が提案されている[3)-7)]．

これまでのほとんどの研究は，対象となる交通ネットワークは単一交差点のみで構成されたネットワークか，せいぜい数個の交差点をもつ小規模ネットワークであり，合流制御が行われる交差点への各方向からの流率は与件として与えられている．コンフリクトを同時に交差点に進入しようとする複数車両間の走行軌跡の交錯として定義したとき，合流制御は，車両の加減速制御などによってコンフリクトの逐次的な解消を目的とする手法である．そのため，十分な制御効果を得るには，車両の加減速制御によって交差点への流入時刻を調整可能なだけの十分な車間距離が存在する必要がある．よって，これらの合流制御手法は，従来の信号機による制御よりも中程度の流入交通量において高い整流効果があると示されているが，交差点への流入率が高い場合には十分な制御効果が得られない場合があることが示されている[8)]．

都市交通網において，ある地区に集中する交通量が多い場合，周辺交差点へ流入する車両数も多くなる．このような状況を想定したとき，交差点の合流制御に加えて，各車両が目的地までの経路における他車両との走行軌跡の交錯が少なくなるような経路配分を行うことで交通流全体の円滑性が向上すると考えられる．しかしながら，都市交通網における経路配分等の既往研究は，交通を流れとして捉えた巨視的視点で解析を行うため，個々の車両の交差点でのコンフリクトを考慮していない．

そこで，本研究では，微視的な交通表現を用いて，経路交錯を減らし，交通流の円滑性が向上するような経路計画手法についての検討を行う．

2. 既往研究と目的

従来の都市交通網における経路計画の課題は，巨視/微視的視点での解析に関わらず，対象とするネットワークや交通量の規模の増加にともない求解にかかる計算量が爆発的に増加する点であった．そこで，近年では，膨大なデータから必要な要素を効率よく抽出するとともに，要素間の複雑な影響関係をモデル化できる機械学習の適用が検討されている [9)-20)].

Nazari et al.[9)]は巡回セールスマン問題としても知られる，複数の目的地をすべて経由する最短経路問題において，事前に計算された最適経路データを学習することで最適経路を幾何的に近似するような経路生成が可能であると示した．Sun et al.[10)]は過去の交通情報データと現在の交通情報に基づく交通状況予測モデルを構築し，負荷分散システムを組み合わせた経路計画アルゴリズムを提案した．また，Basso et al.[11)]は電気自動車を用いた配車サービスにおいて，確率的に発生する配車リクエストに対応し，かつ走行による電力消費を考慮した配車計画問題を強化学習によって経験的に解く手法を提案した．

上述の経路割り当て手法に対して，出発時点で経路を確定させるのではなく，走行中の車両に交通状況に応じて新たな経路を指示する手法や，経路を設定せず目的地までの各交差点での車両の進行方向（流入先のリンク）を選択するモデルを強化学習により構築する手法が存在する．Saunza et al.[12)]は，ネットワーク内の車両の位置情報を用いて k 近傍法により渋滞地点の特定を行い，渋滞地点を通る経路予定していた車両に対して，代替経路を提案する手法を提案した．Yan et al[13)]は，都市交通において交通事故の発生とそれに伴う渋滞状況を想定し，緊急車両の事故地点までの旅行時間が最小となる逐次的な経路選択モデルを Deep Q-Network によって強化学習によって構築した．また，Koh et al.[14)]は車両の逐次的経路選択モデルの強化学習において，目的地までのユークリッド距離など地理的情報を変数に加えることで学習の精度が向上することを示した．

以上のように，様々な経路計画手法において機械学習の有用性が示されてきた．本研究では，前述のようにコンフリクトの削減による交通流の整流化可能性について調査するとともに，強化学習を用いた経路計画システムの構築方法についても検討した．対象交差点に流入する車両に対して，車両の目的地に応じて判断される最適な経路を指示する経路計画システムを車両行動の微視的な表現が可能なメソスコピック交通流シミュレータ上に構築し，システムが使用する経路計画モデルを強化学習によって構築した．

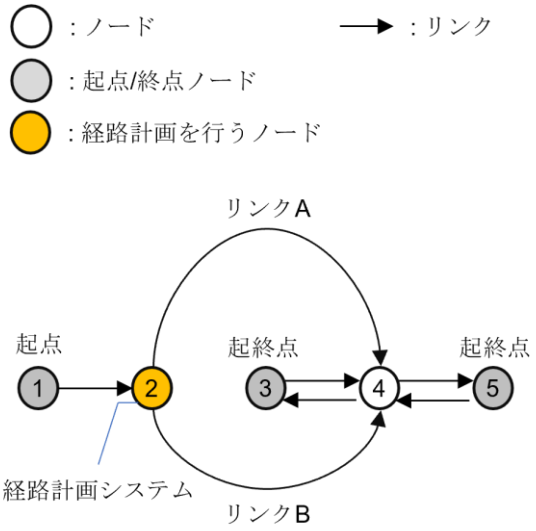


図-1　交通ネットワーク

3. 協調的経路計画とネットワークの概要

本研究では，図-1 に示す交通ネットワークを用いて交差点でのコンフリクトを削減し，車両の旅行時間を最小化するような協調的経路計画システムの構築を行った．

ネットワークにおける起点ノードをノード 1，3，5．終点ノードを 3，5 とした OD 交通流を発生させる時，ノード 2 に仮想的に設置された経路計画システムは，ノード 1 から発生する交通流を制御する．ノード 2 に流入した車両は経路計画システムに目的地を送信し，通信を受けとったシステムは，ネットワーク全体の交通状況を考慮した経路（リンク）A/B へ車両を割り当てる．車両はシステムの指示に従って経路を走行し，ノード 4 を経由して目的地のノード方向のリンクへ流出していく．また，図-2 はノード 4 で表現される交差点において，経路 A/B から交差点に流入しノード 3，5 へ流出する交通流がノ

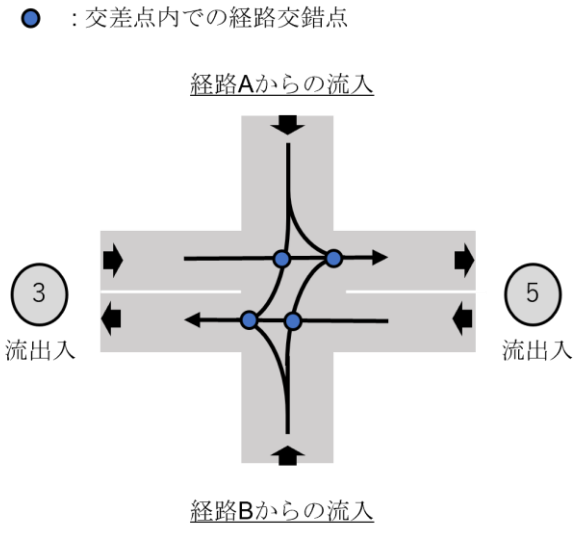


図-2　ノード 4 における交差点内の走行経路

ード 3，5 間の双方向の直進交通流との交差点内での経路の交錯を起こす点を示している．

例えば，ノード 1 から発生してノード 5 へ向かう交通流は，経路 B から交差点に流入し右折するよりも，経路 A から流入し左折をする方が，ノード 5 からノード 3 へ向かう直進交通流との経路の交錯を回避できるため，経路の交錯箇所が少なくなり経路 B から流入するよりも小さな遅れで流入できる可能性が高いと考えられる．しかし，経路 A の距離を経路 B よりも長く設定した場合，移動距離は長いが経路の交錯が少ない経路 A か，移動距離は短いが経路の交錯が多い経路 B のどちらがより早い経路であるかは，それぞれの経路の所要時間や交差点での交錯による遅れ時間の有無や大小によって，状況ごとに変化すると考えられる． ただし，本研究では後続の交通流の円滑性を考慮して，車両が旅行時間が最小でない経路に割り当てられることを許容する動的システム最適計画モデルを構築する．

(1) シミュレータの概要

経路割り当て器の構築と，離散的に車両を表現可能なメソスコピック交通流シミュレータを構築した.

a) 車両速度と位置の更新方法

車両の速度はリンク流出入時のみ更新されるとし，車両がリンクに流入した時刻tにおけるリンク内交通流の速度v_tを対象リンクに流入する車両の速度とする．速度v_tは，以下の Greenshields[14]の式によって計算する．

$$v_t = v_f\left(1 - \frac{k_t}{k_j}\right) \tag{1}$$

ここで，v_f はリンクの最大速度，k_tは時刻 t における交通密度，k_j はリンクの飽和交通密度である．本研究では，すべてのリンクで$v_f = 50[km/h]$,$k_f = 125$[台/km]とした．また，車両の位置は一タイムステップを 1 秒として,各タイムステップごとに更新される．

b) リンク移行ルール

車両がリンク終端へ到達した場合，次に示すリンク移行ルールに従いリンク移行可能であれば，到達したタイムステップでリンク移行を行う．

本研究では，すべての車両が CAVs であり，信号に代わる交差点制御が実装されている環境を想定している．そのため，交差点内での走行軌跡の交錯する複数の車両が同時に交差点に進入しようとした時において，衝突を防ぐため，優先・非優先ルール，もしくは First-Come First-Served (FCFS) ルールによってリンク移行の可否を判断する．より高度な交差点制御の適用については今後の課題とする．

優先・非優先ルールは，現在でも生活道路から幹線道路への流入の制御などで用いられている．各リンクに設定された優先度を基準に，より優先度の高いリンクに存在する車両が優先してリンク移行を可能とする．車両は，リンク終端において自車より優先度の高い車両が，交差点流入部に到達するまでに 3 秒以上のギャップがある場合に交差点へ進入可能とし，そうでない場合はリンク終端で待機する．ただし，対向リンクとの交錯など，車両の優先度が同じ場合，直進＞左折＞右折の順に優先度が高いものとする． また，経路（リンク）A/B を非優先，ノード 3−5 間のリンクを優先に設定した．

FCFS ルールでは，リンクの優先度に関わらず，先着順でリンク移行が可能とする．つまり交差点への進入待機時間がより小さい車両が優先される．ただし，同時刻に待機時間が等しい複数の車両が進入しようとした時は，優先・非優先ルールに従う．

c) Vertical queue による待ち行列の表現

各車両は離散的に表現されるが，物理的な渋滞の長さは考慮せず，リンク終端における車両の待機列は縦積み行列（vertical queue）を用いて表現される．新たに待機列に加わる車両は，縦積み行列の最上部に追加され，最下部に位置する車両から順にリンク流出が行われる．また，待機列内での順序の入れ替え（追い越し）は行わず，一台の車両がリンク流出に必要とする時間を 3 秒間とする．つまり，リンクの最大流出率は一台/3 秒である．

(2) 動的システム最適計画モデルの構築

Actor-Critic は強化学習における代表的な手法の一つであり，エージェントの方策を決定する方策器（Actor）とその方策を評価する批評器（Critic）を用いてエージェントが得られる報酬を最大化するように学習する機械学習手法である．**図-3** に Actor-Critic によって構築される経路割り当て器の構造を示す．今回は対象となる一起点二終点交通流の内，終点ノード別の一起点一終点交通流ごとに経路割り当てを行うエージェントを設定し， 同一環境に存在する二つのエージェントを同時に学習するマルチエージェント学習を行った．

システム最適な交通状況を目指し，交通流の総旅行時間を最小化することを目的に学習を行うため，二つのエージェントが，協調的に行動する必要がある．協調型のマルチエージェント学習では，すべてのエージェントが受け取る報酬を同じに設定することが一般的である．また，Actor-Critic を用いたマルチエージェン学習の代表的な手法 [15]では，複数の方策器がそれぞれ選択した方策の影響を批評器間で共有する構造になっている．既往手法を参考にし，二つの方策器が一つの批評器を共有し，かつそれぞれの交通流の制御から得られる報酬を合算して

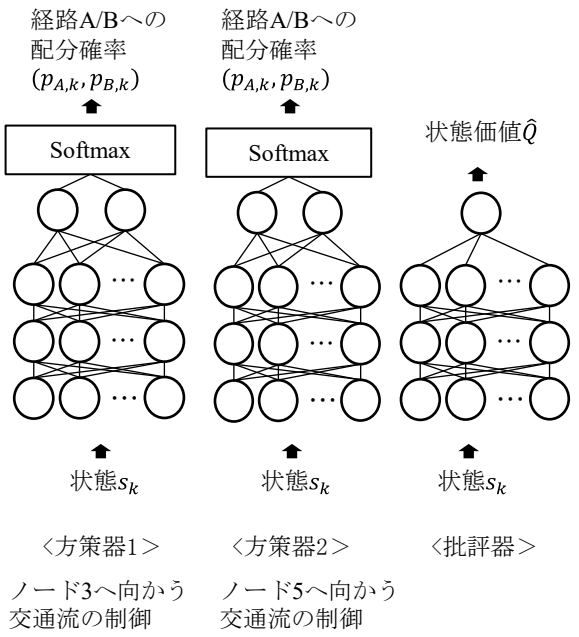


図-3 方策器と批評器の概要

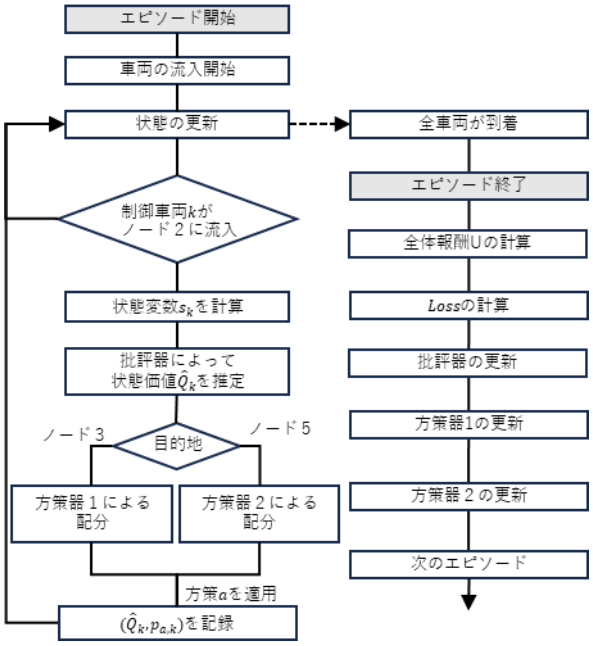


図-4 1エピソードにおける学習の流れ

与えることで，エージェント間の協調行動の学習を試みた．

a) 学習の流れ

ネットワーク内に車両が流入する前の状態を初期状態とし，車両が流入してからすべての車両がネットワークから流出するまでを 1 回のシミュレーション（エピソード）とする．**図-4** は 1 エピソードにおける経路割り当て制御とモデルパラメータの更新の流れを示す．

ノード 1 から発生した車両kが，ノード 2 において経路を割り当てられる際に，その時の交通状態変数s_kをもとに方策選択確率$p_{1,k}, p_{2,k}$と状態価値$\hat{Q}_{s_k}$を以下の式で推定する

$$p_{i,k} = \frac{e^{\pi_i(s_k)}}{\sum_i^n e^{\pi_i(s_k)}} \tag{2}$$

$$\hat{Q}_{s_k} = \theta(s_k) \tag{3}$$

ここで，関数πは方策価値推定関数であり，θは状態価値推定関数でありニューラルネットワークによって近似される．経路計画システムは，方策器から得られた方策選択確率$p_{1,k}, p_{2,k}$となるようにランダムに方策aを決定する(a=1：経路 A へ割り当て，a=2：経路 B へ割り当て)．また，同時にその時の交通状態s_k，採用した方策aの選択確率$p_{a,k}$，批評器による推定状態価値$\hat{Q}_k$を記録する．ただし，方策器による経路割り当ては 1 秒から 360 秒までの 6 分間の間にネットワークに流入した車両を対象とし，それ以外の車両はリンクの所要時間が短い経路に割り当てる．

b) モデルパラメータの更新

強化学習では，上述のようにエージェントによる経路割り当て制御と状態価値推定を行いながらシミュレーションを実行し，得られた報酬とモデルが予測した報酬の誤差が小さくなるようにモデル（方策器/批評器）のパラメータの更新を行う．実際に経路計画を行った車両台数をnとした時，モデルがシミュレーション終了時に得られる全体報酬は以下の式で計算される．

$$U = \frac{1}{100}\left(c - \sum_{k=1}^{k=n}(t_k + d_k \cdot w)\right) \tag{4}$$

ここで，t_kはk番目に制御された車両の自由走行時間，d_kはコンフリクトに起因する停止時間，wは停止時間に与える重み，cは任意の定数である．ただし，今回は$w = 1$とし，$(t_k + d_k)$で表される車両kの旅行時間（実走行時間）から報酬を計算した．

k番目の車両の割り当てによって得られる報酬R_kは全体報酬Uをもとに次のように与える．

$$\begin{aligned} R_k &= U, \qquad k = n \\ R_{k-1} &= R_k \cdot \gamma^{dt_{k-1,k}}, \quad 2 < k < n \end{aligned} \tag{5}$$

$\gamma(= 0.99)$は時間割引率，$dt_{k-1,k}$は$k-1$番目とk番目の車両の制御の時間間隔である．本研究では，一般的な強化学習タスクはエージェントが一定間隔で行動することを前提としている．しかし本研究では，車両の流入ごとにエージェント（経路割り当て器）は行動を行うため，

行動ごとの時間間隔による報酬の割引を$\gamma^{dt_{k-1,k}}$によって表現する．

Actor-Critic では，批評器によって推定される状態価値$\hat{Q}$を基準に方策器を更新していく．k番目の車両に経路割り当てを行う時の状態s_k，方策器によって採用された方策iの選択確率$p_{i,k}$とする．批評器による推定状態価値$\hat{Q}_k$とすると，真の状態価値Q_kは全体報酬から逆算した報酬R_kに等しい．この時，方策器と批評器の推定誤差$Loss_{actor}$，$Loss_{critic}$は以下の式で計算される．

$$\delta_k = \hat{Q}_k - Q_k \tag{6}$$

$$Loss_{critic} = \begin{cases} \frac{1}{2}\delta_k^2, |\delta_k| < 1 \\ \delta_k - \frac{1}{2}, |\delta_k| > 1 \end{cases} \tag{7}$$

$$Loss_{actor} = -\log(p_{a,k})\,\delta_k \tag{8}$$

この誤差が小さくなるよう，誤差伝播法を用いてモデルのパラメータを更新する．ただし，方策器 1，2 はそれぞれの方策器を用いて行った制御から得られる誤差のみを用いるものとする．誤差の計算には全体報酬が必要となるため，モデルのパラメータは 1 回のエピソード終了毎に，そのエピソードで行ったn回の行動から得られた誤差すべてを用いて行う．

c) 状態変数の設定

方策器/批評器に入力する状態sは，交通密度[台/m]，平均速度[km/h]，ノード 3 へ向かう車両台数，ノード 4 へ向かう車両台数の 4 指標をすべてのリンクにおいて，始端から終端まで 10 [m]間隔ごとに集計した値で構成される行列である．

表-1　交通条件

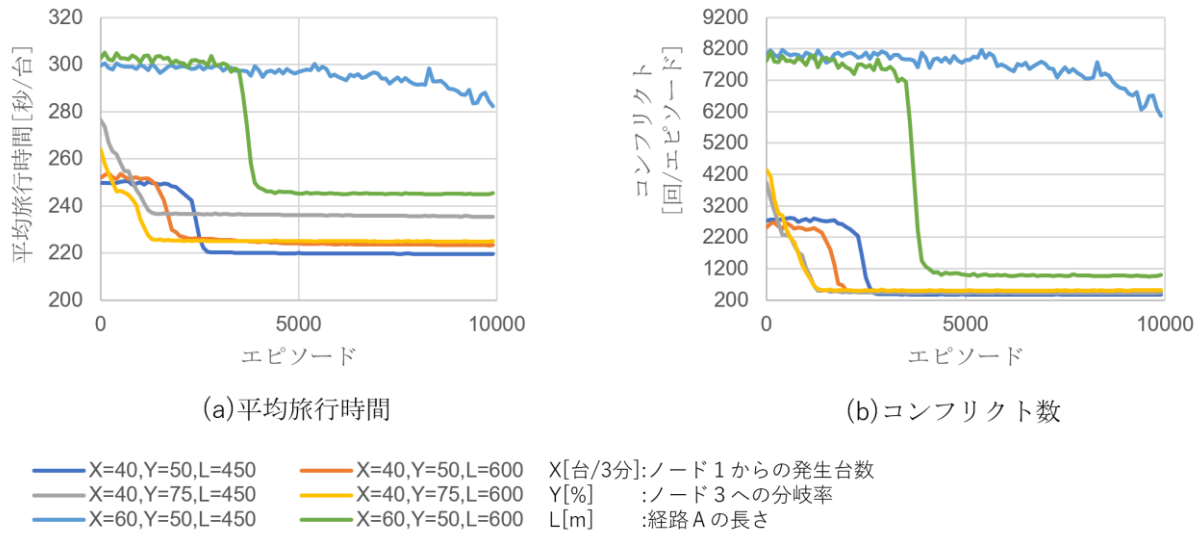
設定	ノード 1 からの発生台数 X [台/3 分]	ノード 3 への分岐率 Y [%]	経路 A の長さ L [m]
1	40	50	450
2	40	50	600
3	40	75	450
4	40	50	450
5	60	50	450
6	60	50	600

(3) 学習の設定

ノード 1 からノードmへ向かう交通量をx_m[台/3min.]とした時，学習時の交通条件の操作変数としてノード 1 からの発生交通量 $X(=x_3+x_5)$[台/3min.]，ノード 3 への分岐率 $Y=x_3/X$ [%] 経路 A の長さをLと定義する．**表-1** に示すように，操作変数の値を変えながら複数の環境において学習を行うことで，交通条件の違いによる経路計画への影響を考察する．ただし各交通条件で共通する設定として，ノード 3 からノード 4 への交通量を 30 [台/3min.]（ノード 4 からノード 3 も同様），経路 B の長さを 300 [m]，それ以外のリンク長を 1,000 [m]とした．また，車両の出発時間は交通条件ごとに事前にランダムに決定した値を固定した状態で与える．

5. 実験結果と考察

各交通条件で 1 万回エピソードを実施し，各エピソードごとの制御対象の交通流の総旅行時間と，総コンフリクト数を記録した．ただし，シミュレーション上でのコンフリクト数は車両が交差点流入を試みた際に自車よりも優先度の高い車両の存在によって交差点流入が妨げら



(a)平均旅行時間　　(b)コンフリクト数

図-5　優先・非優先ルールでの学習結果

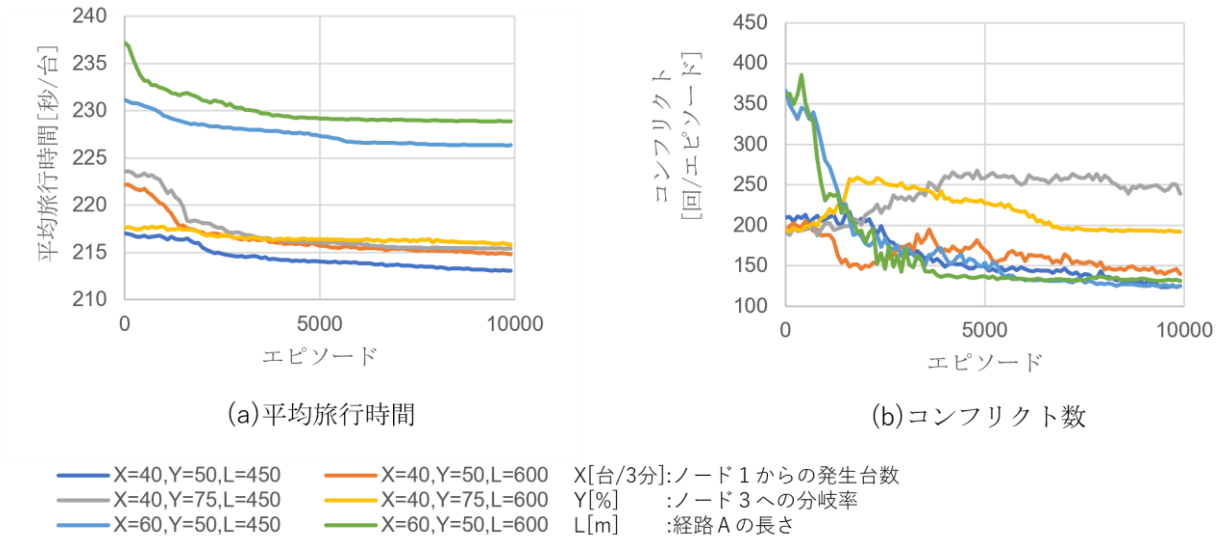


図-6　First-Come First-Served ルールでの学習結果

れ，待機されられた回数である．

(1) 平均旅行時間とコンフリクト回数

図-5 は，優先・非優先ルールにおける，学習過程での 100 エピソードごとの平均旅行時間 [秒/台]とコンフリクト回数の変化を示している．優先・非優先ルールによって交差点制御を行う学習環境では，ノード 1 からの発生交通量X によって平均旅行時間が大きく異なる．これは，ノード 2 におけるリンク移行が非優先リンクから優先リンクへの移行であるため，非優先リンクでの待機列が発生しやすく，発生交通量の影響が大きいためである．ほとんどの環境設定で 1 万エピソード経過時点で学習初期よりも平均旅行時間が減少した状態への収束が示された．また，平均旅行時間の収束がみられない，X=60 [台/3 分], ノード 5 への分岐率Y＝50 [%]，経路 A の長さL=450 [m] の環境でも，1 万エピソード経過時点で平均旅行時間が減少傾向にあり，追加の学習を行うことで，他の環境と同様に収束可能であると期待できる．

学習過程における平均旅行時間とコンフリクト数の推移はほぼ同様の傾向であることが分かる．このことから，優先・非優先ルールで制御される交差点を通過する交通流ではコンフリクトを減らすような経路計画を行うことで，交通流の円滑性を向上できることが示された．

図-6 は，FCFS ルール環境での学習結果を示している．優先・非優先ルールと同様に，ほとんどの交通条件で学習が進むに従い，平均旅行時間とコンフリクトが減少することが示された．

一方で，X=40 [台/3 分]，ノード 5 への分岐率Y＝75 [%] の 2 パターンの学習環境で，1 万エピソード経過時点での，1 エピソードにおける平均的なコンフリクト数は学習初期と同程度かそれよりも高い結果となった．

表-2　優先・非優先での経路計画結果

X, Y, L の設定	ノード 5 への交通		ノード 3 への交通	
	経路 A への割り当て率 [%]	経路 B への割り当て率[%]	経路 A への割り当て率[%]	経路 B への割り当て率[%]
40, 50, 450	92.5	7.50	0.00	100
40, 50, 600	62.5	37.5	0.00	100
40, 75, 450	100	0.0	0.00	100
40, 75, 600	91.7	8.30	0.00	100
60, 50, 450	95.0	5.00	5.00	95.0
60, 50, 600	100	0.00	0.00	100

X[台/3 分] : ノード 1 からの発生台数

Y[%] : ノード 3 への分岐率

L[m] : 経路 A の長さ

FCFS ルールでは先着順に流入可能であるため，一度コンフリクトによって待機状態になっても，せいぜい数秒経過すれば，コンフリクトせずに交差点への流入が可能となる．両ルールでの学習時のコンフリクト数を比較したとき，学習初期段階から FCFS ルール環境は優先・非優先ルール環境よりもコンフリクト数が大幅少ない．今回の一部の交通量設定では，FCFS ルールで制御される交差点でコンフリクトによる混雑が発生しづらい程度の入力交通量であり，コンフリクトによる遅れが交通流に与える影響は小さかったと考えられる．

(2) 経路割り当て結果の分析

表-2 は優先・非優先ルールで学習した経路割り当て器を用いて，学習時と同じ条件で経路計画を行った時の，各経路に割り当てられた車両の割合を終点ノードごとに集計した結果を示す．ノード 1 を起点とし，ノード 3 へ向かう交通流の経路計画では，ほぼすべての環境設定において，経路 B への割り当て率が 100%であった．これ

表-3　First-Come First-served での経路計画結果

X, Y, L の設定	ノード 5 への交通		ノード 3 への交通	
	経路 A への割り当て率[%]	経路 B への割り当て率[%]	経路 A への割り当て率[%]	経路 B への割り当て率[%]
40, 50, 450	40.0	60.0	0.0	100
40, 50, 600	30.0	70.0	0.0	100
40, 75, 450	0.00	100	0.0	100
40, 75, 600	0.00	100	10.0	90.0
60, 50, 450	53.3	46.7	0.0	100.0
60, 50, 600	43.3	56.7	0.0	100.0

X [台/3 分]：ノード 1 からの発生台数

Y [%]：ノード 3 への分岐率

L [m]：経路 A の長さ

は，3 章で述べたように，ノード 3 へ向かう交通流は経路 B から流入することで，ノード 3 からノード 5 へ向かう交通流とのコンフリクトを回避することが出来るためであると考えられる．同様の理由で，ノード 5 へ向かう交通流では移動距離は長いがコンフリクトの起こりにくい経路 A への割り当てが多くなる結果となったと考えられる．

表-3 は，FCFS における経路計画の結果を示している．優先・非優先ルールでの結果と同様にノード 3 へ向かう交通流は，ほとんど経路 B に割り当てられる結果となった．一方で，ノード 5 へ向かう交通流での結果は優先・非優先の結果と異なった．例えば，ノード 1 からの発生台数X=40 [台/3 分]，ノード 3，ノード 5 への分岐率Y=50 [%]，経路 A の長さL=450 [m]の時，経路 B への割り当て率は 6 割と経路 A への割り当て率より高い結果となった．

前述のように FCFS ルール環境では，優先・非優先ルールと比較して，交差点でのコンフリクトによる混雑が発生しづらい．そのため．コンフリクトの発生しやすさよりも移動距離を考慮してより短い経路 B への割り当てが増加したと考えられる．これはX=40 [台/3 分]，X=60 [台/3 分]の割り当て結果に着目したとき，経路 A の長さLを 450 [m]から 600 [m]にして経路 A の通過に必要な所要時間を増加させた場合に，経路 B への割り当てが増加することからも，経路の所要時間を考慮した結果であったといえる．

6.　結論と今後の展望

本研究では，強化学習によって経路計画システムの構築を行い，交差点でのコンフリクトが交通流に与える影響を分析した．対象交通流の発生台数や，終点への分岐率を変化させながら学習を行ったところ，コンフリクトによる旅行時間の遅れの影響が大きい優先・非優先ルールでは，目的地に応じて交差点でのコンフリクトが発生しづらい経路への割り当てを行うことで，コンフリクトの発生数を削減し，旅行時間を減少させることが可能であると示された．一方で，コンフリクトによる交通流への遅れの影響が小さい First-Come First-Served ルール交差点制御の環境では，一部の条件設定でコンフリクト数を考慮せず経路の所要時間に応じた経路計割り当てを行うことで旅行時間を減らすように学習が行われた．いずれにしても，ほとんどの条件で経路計画によるコンフリクト数の削減とそれに伴う旅行時間の削減が Actor-Critic を用いた強化学習で可能であることが示された．

しかしながら，本研究で学習の対象にした交通流は一起点二終点とシンプルであるのに加えて，割り当て可能な経路も二経路であった．また，学習時のエピソードごとの車両の出発時間が固定であり，学習した経路割り当て器の汎用性は低い．そのため，車両の出発時刻をランダムにした条件や中規模ネットワークで学習が可能であるか調査が必要である．さらに，一起点二終点の交通流に対して，一起点一終点ごとに方策器を設定し学習を行ったが，実際には，終点（目的地）は無数に存在するため，終点ごとに方策器を学習するのは現実的でない．車両ごとの終点を入力変数として与え，複数の目的地を持つ車両が混在する交通流を一つの方策器で割り当てが可能となるようなモデルの学習についても検討したい．

謝辞：本研究は JSPS 科研費 19H02260 の助成と JST 及び名古屋大学による名古屋大学融合フロンティアフェローシップの支援を受けて得た成果の一部です．ここに記して謝意を表します

REFERENCES

1) Dresner, K. and Stone, P.: Multiagent traffic management: a reservation-based intersection control mechanism, Proc. AAMAS, New York, NY, USA, pp. 530-537, 2004.
2) Xu, B. *et al.*, “Distributed conflict-free cooperation for multiple connected vehicles at unsignalized intersections,” *Transp. Res. C*, vol. 93, pp. 322-334, 2018, DOI: 10.1016/j.trc.2018.06.004
3) Lee, J. and Park, B.: Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment, IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 1, pp. 81-90, 2012.
4) Kamal, M.A.S., Imura, J., Hayakawa, T., Ohata, A. and Aihara, K.: A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights, IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 3, pp. 1136-1147, 2015.
5) Qian, X. *et al.*, Decentralized model predictive control for smooth coordination of automated vehicles at intersection, *2015 Eur. Control Conf. (ECC)*, vol. 2015, 2015, pp. 3452-3458, DOI: 10.1109/ECC.2015.7331068
6) Fayazi, S. A. and Vahidi, A. Vehicle-in-the-loop (VIL)

verification of a smart city intersection control scheme for autonomous vehicles, *2017 IEEE Conf. Control Technol. Appl. (CCTA)*, Maui, HI, USA, 2017, pp. 1575-1580, DOI: 10.1109/CCTA.2017.8062681
7) Guo, Z., Wu, Y., Wang, L. and Zhang, J. Coordination for Connected and Automated Vehicles At Non-Signalized Intersections: A Value Decomposition-Based Multiagent Deep Reinforcement Learning Approach, in *IEEE Trans. Veh. Technol.*, 2022, DOI: 10.1109/TVT.2022.3219428.
8) 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
9) Mohammadreza N, Oroojlooy,A., Lawrence V. S., Martin T, Reinforcement Learning for Solving the Vehicle Routing Problem, arXiv:1802.04240,2018
10) N. Sun, H. Shi, G. Han, B. Wang and L. Shu, "Dynamic Path Planning Algorithms With Load Balancing Based on Data Prediction for Smart Transportation Systems," in IEEE Access, vol. 8, pp. 15907-15922, 2020, doi: 10.1109/ACCESS.2020.2966995.
11) Li, Q.; Xu, Y.; Bu, S.; Yang, J. Smart Vehicle Path Planning Based on Modified PRM Algorithm. Sensors 2022, 22, 6581. https://doi.org/10.3390/s22176581
12) Younes,R., Najem M., A real-time path planning for reducing vehicles traveling time in cooperative-intelligent transportation systems,Simulation Modelling Practice and Theory,Volume 123,2023,102710,ISSN 1569-190X,https://doi.org/10.1016/j.simpat.2022.102710.
13) Longhao,Y.,Ping,W.,Jingwen,Y.,Hu,Y.,Han,Y.Yao,J.,Refined Path Planning for Emergency Rescue Vehicles on Congested Urban Arterial Roads via Reinforcement Learning Approach, Journal of Advanced Transportation,2021,doi 10.1155/2021/8772688
14) Koh, S. et al.,Real-time deep reinforcement learning based vehicle navigation,Applied Soft Computing,Volume 96,2020,106694,ISSN 1568-4946,https://doi.org/10.1016/j.asoc.2020.106694.
15) Greenshiedlds, B. D: A study in Highway Capacity,H. R. B, Proceedings, vol.14, p.448, 1935
16) Ryan,L., Yi,W., Aviv,T., Jean,H., Pieter,A., Igor,M., Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments, arXiv:1706.02275,2017

(Received July 1, 2022)
(Accepted November 1, 2022)

PRELIMINARY STUDY ON COLLABORATIVE ROUTE ALLOCATION REINFORCEMENT LEARNING FORCUSING ON INTERSECTION CONGESTION

Shintaro KATAGORO, Tomio MIWA and Takayuki MORIKAWA