

# 強化学習型オンランプ制御手法に関する研究\*

## Research on Reinforcement Learning Ramp Metering\*

王興拳\*\* 宮城俊彦\*\*\*高木朗義\*\*\*\*

By Xingju WANG\*\*· Toshihiko MIYAGI\*\*\*· Akiyoshi TAKAGI

### 1. はじめに

高速道路における交通渋滞の発生は、道路運行の効率低下、エネルギー消費と環境汚染の増加をもたらすため、これまでも渋滞を緩和するための各種の交通制御が考えられてきた<sup>1,2,3)</sup>。

渋滞は渋滞箇所と渋滞延長が時々刻々と変化するダイナミックな現象を示すため、制御モデルを作成する前段階として、交通流の動的な変化をシミュレートするモデルが必要になる。王・宮城<sup>4)</sup>はTreiber、HelbingらのIDM (Intelligent Driver Model) <sup>5)</sup>に独自の車線変更モデルを加えたシミュレーションモデルを作成し、Helbingが指摘するlocalized clusters、Homogeneous congested trafficあるいはKerner<sup>6)</sup>が指摘する相転移現象を再現した。また、シミュレーションにおけるパラメータに対する感度分析結果から、本線とオンランプの車両流入量が渋滞発生の基本要因になることを明らかにしている。オンランプ区間で交通流入量の有効な制御ができた場合、交通混雑を事前に防ぎ、交通流を円滑化することができる。

本研究の目的は、人工知能における強化学習モデルを導入し、動的ミクロ交通流シミュレーションと組み合わせることによって効果的なオンランプ区間の制御手法を提案することである。

### 2. オンランプの制御手法

オンランプ制御策は、高速道路と高速道路ネットワー

クにおける効率的な輸送管理において重要な手段である。Papageorgiou<sup>8)</sup>によれば、オンランプ制御手法は以下のように大別される。

- (1) 実時間の測定値を用いた反応型制御手法
- (2) 交通需要と予測モデルを基礎とし、高速道路ネットワークにおける最適な交通状態を長時間に亘って維持する先行型制御手法

また、高速道路ネットワーク全体の最適化において、反応型制御手法を居所的に利用する組み合わせ制御手法も可能である。

反応型制御手法は局所型と連携型に分けられる。居所的反応型制御は、個々のランプ交通量の測定値を用いて制御を更新する手法である。一方、連携型反応型制御は、本線道路の交通量の測定値を用いて制御を更新する手法である。一般に局所的反応型制御の方が運用面からは利用しやすいといえるが、交通渋滞状態下では、連携型反応型制御のほうが局所的反応型制御に比べて、有効だと証明されている<sup>9)</sup>。

よく知られた局所反応型制御手法には、DC (Demand-Capacity ; 需要-容量)、OCC (Occupancy ; 占有率)<sup>10)</sup>、ALNEA<sup>11)</sup>などがある。DCとOCCはそれぞれ上流交通需要-容量そして本線占有の測定値に基づいてパラメータを設定する制御手法である。ALNEAは本線交通量の測定値に基づいて、オンランプの占有率を設定する制御手法である。ALINEAは欧州におけるいくつかの国で適用例があり<sup>12)</sup>、DCとOCCに比較して有効性が高いとされている<sup>9)</sup>。しかし、オンランプ付近の交通状態によって制御が決定されるため、高速道路の下流に生じた渋滞を考慮することができないという問題がある<sup>13)</sup>。この問題を解決するために、岩田・坪田・川嶋<sup>13)</sup>は交通シミュレータによる予測値を利用したランプ制御手法を提案しているが、いくつかの問題点が残されている。まず、交通流シミュレータのVISSIM自身に課題がある。例えば、VISSIMでは車両の走行特性に応じた内部パラメータの調整を必要とするが、技術資料が公表されていないため、日本の車両特性に応じたパラメータ調整が難しい。

\*キーワード: 強化学習、オンランプ制御、交通流シミュレーション

\*\*地域修士、岐阜大学工学研究科

\*\*\*正員、工博、東北大学情報科学研究科 教授

(仙台市青葉区荒巻字青葉 6-6、TEL:022-795-7495)

\*\*\*\*正員、工博、岐阜大学工学研究科 教授

(岐阜市柳戸1-1、TEL:058-293-3307)

また、一般のユーザーではVISSIMにユーザー独自の制御手法や機能を追加することはできない。また、岩田・坪田・川嶋のランプ制御手法では、ランプの開閉パターンのみが考えられており、オンランプの流入量そのものを制御対象にしているわけではない。さらに、シミュレーションを用いて単位時間内のシミュレータによる予測を用いたランプ制御手法を適用しているが、渋滞交通量および長時間の最適化などを考慮しなければならない、あるいは、最適コントロール関数が必要である。

### 3. 強化学習型オンランプ制御手法

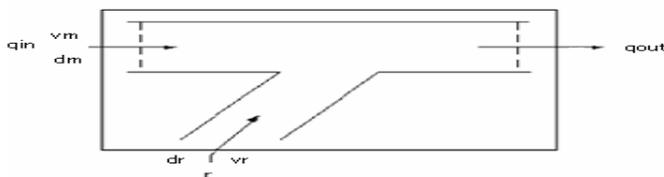
強化学習(Reinforcement Learning)とは、ある環境下におけるエージェントが、現在の状態を観測し、取るべき行動を決定する問題を扱う機械学習の一種である<sup>14)</sup>。エージェントは行動を選択することで環境から報酬を得る。強化学習は一連の行動を通じて報酬が最も多く得られるような方策(policy)を学習する。未知の学習領域を開拓していく行動と、既知の学習領域を利用する行動とをバランス良く選択することができる。

強化学習では、エージェントは様々な行動を試し、より多く報酬が得られる行動を選択する。他の学習則と異なる特徴は、正しい行動を教えられるのではなく、取った行動を評価することで学習することである。そのため、二つのプロセスを繰り返す。一つのプロセスは、現在のポリシーに従って行動し、得られた報酬から価値関数(Value function)を作成すること(方策評価)であり、もう一つのプロセスは、現在の価値関数から最適な方策を探索すること(方策改善)である。以下、強化学習の一手法であるQ学習について説明する。

強化学習における最も重要な発展形の一つであるQ学習は方策オフ型TD制御である。最も簡単な形式は1ステップQ学習と呼ばれ、次のように定義される。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + a [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (1)$$

式(1)において、学習で獲得される行動価値関数Qは、使用されている方策とは独立にQ\* (最適行動価値関数)を直接近似する。この手法では、方策とは別に状態行動変数が更新され、ステップサイズ・パラメータに関する確率近似の条件下で、確率1でQ<sub>t</sub>はQ\*に収束することが示されている。



### 図-2 強化学習型オンランプ制御手法

オンランプ制御手法の問題を解決するために、強化学習型オンランプTD制御手法を提案する。図-2に示すように、 $qin$ : 本線上流交通量、 $r$ : ランプ流入交通量、 $qout$ : 本線下流交通量、 $dm$ : 本線上流密度、 $dr$ : ランプ密度、 $vm$ : 本線上流平均速度、 $vr$ : ランプ平均速度である。

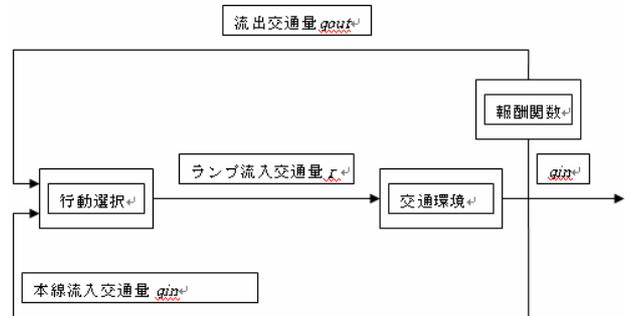


図-3 強化学習型オンランプ制御手法の概念図

図-3にしたがって、強化学習オンランプ制御手法を構成する基本要素は以下のようである。

#### (1) 方策 (policy)

最適化条件は以下に示すとおりである。

- I 本線で渋滞が発生しない。
- II ランプで渋滞が発生しない。
- III 下流交通量は最大になる。

IとIIについては、王・宮城<sup>15)</sup>によるシミュレーション結果より、本線速度は20m/h以下、密度15台以上になると大渋滞へ転移する可能性が大きいことがわかっているので、IとIIについては、次の条件をおく。

$$\begin{aligned} dm < 15 & \quad dr < 15 \\ vm < 20 & \quad vr < 10 \end{aligned}$$

#### (2) 報酬関数 (reward function)

ランプ流入交通量  $r$

#### (3) 価値関数 (value function)

総流出交通量  $Q$

#### (4) 環境モデル (model)

交通環境 (本線流入交通量、滞留交通量、ランプ流入交通量)

式(1)を次のように書き換える。

$$Q(qin_t, r_t) \leftarrow Q(qin_t, r_t) + a [qout + \gamma \max_a Q(qin_{t+1}, r) - Q(qin_t, r_t)] \quad (2)$$

区間内に滞留する交通量 $q$ は次式で与えられる。

$$q = qin + r - qout \quad (3)$$

滞留交通量 $q$ を考慮して、上流交通量を次式で更新する。

$$qin_{t+1} \leftarrow qin_{t+1} + q \quad (4)$$

上流交通量 $q_{in}$ は観測によって得られるものとし、これを既知量（状態変数）として扱う。ランプ流入交通量 $r$ は制御量として、選択行動変数と置く。また、 $q_{out}$ は行動選択に伴う報酬である。シミュレーションを用いて、予測結果によって、最適な $Q$ を決める。図-4に $Q$ 学習のアルゴリズムを示す。

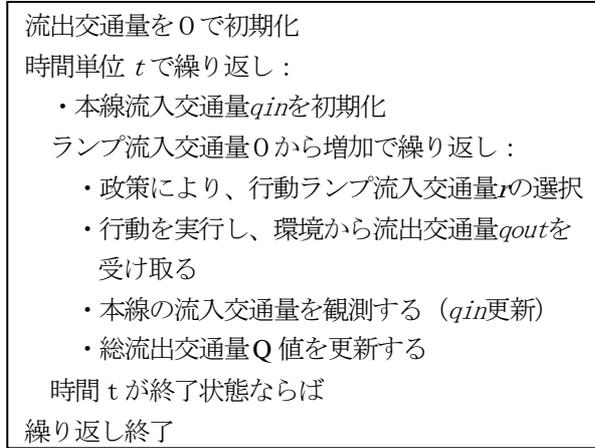


図-4 TD制御アルゴリズム

強化学習型オンランプTD制御の手順を図-5および以下に示す。

- (1) 本線上流交通量の入力
- (2) 行動（ランプ交通流入量）を選択して、シミュレーションを行う
- (3) シミュレーションの結果によって、最適な制御行動の選択
- (4) 総流出交通量 $Q$ と本線上流流入交通量の更新
- (5) 時間単位で繰り返し
- (6) 時間終了ならば、繰り返しの終了

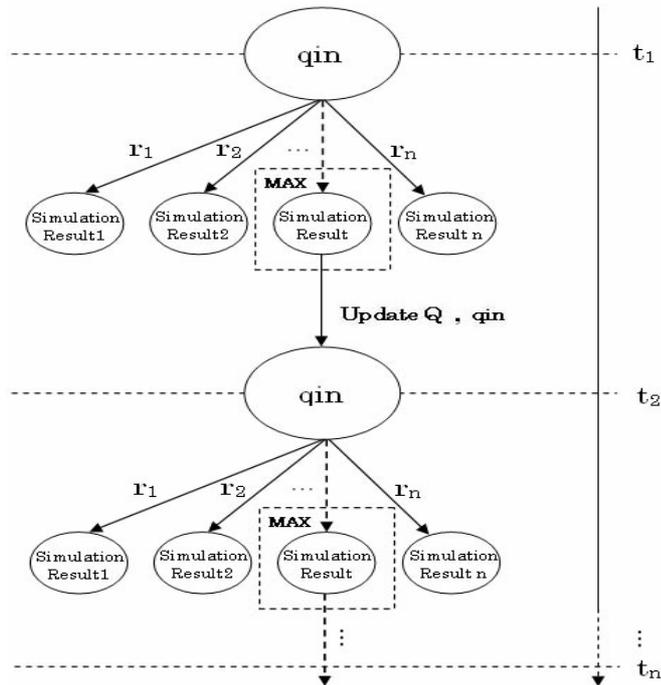


図-5 TDランプ制御の手順

#### 4. オンランプ制御手法の検証

本研究におけるシミュレーションでは、道路の合流部と観測点の構造を図-6に示すように設定した。ここで、本線道路の長さは1,000m、ランプの長さは200m、ランプの合流部の長さは100m、ランプと本線道路との合流角度は30度と設定した。観測点の距離は基準点からの距離で表している。

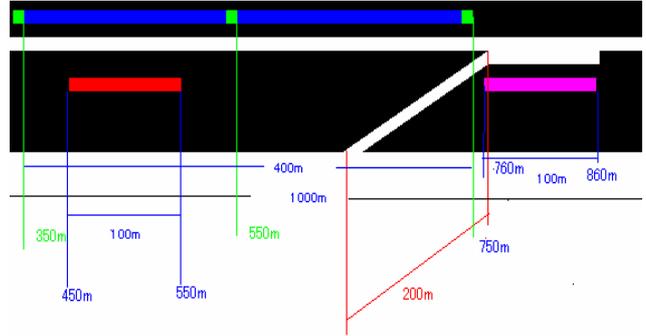


図-6 シミュレーションで用いた道路合流部の構造

ケース	時間軸 (分)	本線交通量 (台/h)	ランプ交通量 (台/h)	
			制御なし	制御あり
A	1.5	600	800	800
	3	1200		600
	4.5	1800		0
	6	2200		600
B	1.5	1200	800	800
	3	1800		600
	4.5	2200		0
	6	1600		600
C	1.5	600	800	800
	3	800		600
	4.5	1000		0
	6	1200		600

表-1 パラメータの設定と最適な制御行動の選択

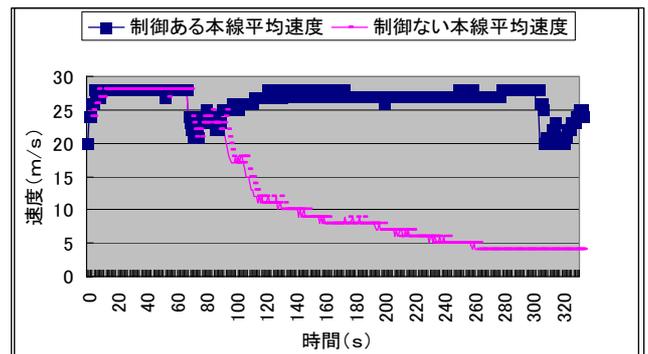


図-7 ケースBには本線における平均速度の時間変化

表-2 通過交通量と平均旅行時間による制御効果

ケース		制御なし	制御あり	変化量	渋滞緩和率
A	通過交通量 (台/h)	1850	1860	10	0.54%
	平均旅行時間 (s/台)	73	40	-33	-45.38%
B	通過交通量 (台/h)	1840	2120	280	15.22%
	平均旅行時間 (s/台)	78	38	-40	-51.28%
C	通過交通量 (台/h)	1560	1680	120	7.69%
	平均旅行時間 (s/台)	38	43	5	10.97%

表-1に示すように、ケースA、BとCを設定した。「制御なし」の場合、ランプ平均交通量は800台/hと設定する。「制御あり」の場合、最適行動は表-1に示す結果となった。ケースBには、渋滞交通量を考慮した場合、本線交通量が増加し、1,600台/h~1,800台/hとなると、流入制御が必要になり、行動(600台/h)を選択するという結果を得る。また、このときの平均速度の時間変化を図-7に示す。「制御なし」で流入を許可し続けると渋滞発生後4m/sまで下がるが、「制御あり」の場合、平均速度は約25m/s付近まで増加させることができる。交通量が2,000台/hを超えるようになってくるとランプを一時的に閉鎖する必要が生じる。

ランプ制御の効果を通過交通量と平均旅行時間によって見た結果を表-2に示す。ケースBには、通過交通量で見ると280台/h増加し、渋滞緩和効果は15.22%となる。また、平均旅行時間で見ると、40s/台減少し、渋滞緩和率は51.28%となった。一方、本線交通量が多いケースAとBは平均旅行時間で見て、渋滞緩和率は45%以上になった。一方、本線交通量少ないケースCには、通過交通量は120台/h増加したが、平均旅行時間にみて、5秒を増加した。本線交通量が多い場合、提案した制御手法の効果は大きいことが明らかにわかった。

## 5. まとめ

オンランプ区間における交通渋滞を緩和するために、ランプ制御手法が有効である。本研究では、時々刻々と

変化する本線上流交通量に対するシミュレーションを実施し、最適行動選択ツールとする強化学習オンランプ制御手法を提案した。検証の結果、交通流を円滑化することが確認された。今後、実際の観測データを利用し、手法の有効性を確認していきたい。

## 参考文献

- 1) 明神証, 坂本破魔雄, 岩本俊輔: 待ち行列を考慮したLP制御, 交通工学, Vol.8, No4, pp.15-22, 1975.
- 2) 松井寛, 佐藤佳朗: 都市高速道路の動的流入制御理論に関する研究, 土木学会論文報告集, No.326, pp.103-114, 1982.
- 3) 秋山孝正, 佐佐木綱: ファジィ流入制御モデルを用いた交通制御方法の評価と検討, 土木学会論文集, 第413号, pp.77-86, 1991.
- 4) 王興挙, 宮城俊彦: 希望速度を考慮したモデルによる渋滞波及現象のシミュレーション, 土木計画学研究・講演集, Vol.32, 120, 2005.
- 5) 王興挙, 宮城俊彦: 渋滞波及現象モデルによるオンランプ区間での混雑現象分析, 交通工学研究発表会論文報告書, Vol.26, pp.29-32, 2006.
- 6) M.Treiber, A.Hennecke and D.Helbing: Congest traffic state in empirical observations and microscopic simulation, Review E. vol.62, pp1805-1824, 2000.
- 7) Kerner, B.S.: The Physics of Traffic, Springer, Germany, 2004.
- 8) M.Papageorgiou, A.Kotsialos: Freeway ramp metering: an overview, IEEC Conference on Intelligent Transportation System 3, pp.271-281, 2002.
- 9) M.Papageorgiou, H.Haj-Salem, F.Middleham: ALINEA Local ramp metering: summary of field results, Transportation Research Record 1603, pp.90-98, 1997.
- 10) D.P.Masher, D.W.Ross, P.J.Wong, P.L.Tuan, H.M.Zeidler, S.Petracek: Guidelines for design and operation of ramp control systems, Stanford Research Institute, Menlo Park, California, 1975.
- 11) M.Papageorgiou, H.Haj-Salem, J-M.Blosseville: ALINEA: A Local feedback control law for on-ramp metering, Transportation Research Record 1320, pp.58-64, 1991.
- 12) E.Kenis, R.Tebeos: Ramp metering Synthesis, Report of the Centrico Project, European Commission, Brussels, Belgium, 2001.
- 13) 岩田理史, 坪田幸政, 川嶋弘尚: 渋高速道路におけるランプ制御に関する比較研究, 交通工学研究発表会論文報告書, Vol.26, pp.73-76, 2006.
- 14) S.S.Richard, G.B.Andrew, 三上貞芳(訳), 皆川雅章(訳): 強化学習, 森北出版, 2000.