7D-01

# マルチエージェント強化学習による交通信号機制御に向けて

岡野拓哉 †‡　大西正輝 †‡　野田五十樹 ‡

† 筑波大学システム情報工学研究科　　‡ 産業技術総合研究所人工知能研究センター

## 1 背景

### 1.1 信号機制御

交通渋滞を緩和するためには信号機を適切に制御することが重要である. しかしながら, 信号機は配置地点, 時間, 状況によって最適制御方策は異なるため, 人手で制御することは難しい. そこで, 現在の信号機周辺の状況から, 信号機の最適制御方策を自動的に獲得するシステムが望まれる.

信号機には複数の制御対象のパラメータが存在する. 大きく分けると, 青から赤までの一連の流れの長さを決定するサイクル長, サイクル内の各現示の比率を決定するスプリット, 隣接した信号機間の青信号の開始時間のずれを決定するオフセットの 3 つの制御対象のパラメータが存在する. 一度にすべてのパラメータの調整を行うことは難しいため, 本研究では, スプリットを制御することで, 交通渋滞を減少させることを目指す. つまり, サイクル内の各現示の比率を調整することで, 交通渋滞を減少させることを目指す.

### 1.2 強化学習とマルチエージェント強化学習

強化学習は試行錯誤を行い最適方策を学習するためのフレームワークである. 強化学習では, 現在の状態 $s$ において, ある行動 $a$ を選択し, その行動を実行した後に変化した状態 $s'$ 及び, 報酬 $r$ を得ることにより, 最適方策を学習する.

Deep Q Learning(DQL)[1] は深層学習と強化学習を組み合わせた学習アルゴリズムである. DQL では, 行動価値関数をニューラルネットワークで関数近似することで, 大規模な状態空間を必要とする複雑な問題においても学習可能なことが知られている.

複数の強化学習エージェントが同じ環境で同時に学習行動を行う場合をマルチエージェント強化学習と呼び, 数多くの研究がなされてきた [2]. マルチエージェント強化学習の最もシンプルな形式は Independent Learners(IL) である. IL では, 各エージェントが各々方策を持ち, 他のエージェントを環境の一部とみなして独立で学習する. IL はエージェント数が増加したとしても, 各エージェントの状態空間の大きさは変化しないため, スケーラビリティに優れている. 協調行動を促す仕組みが必要等の問題点もあるが, IL は最も実用的な形式である. IL の中でも, すべてのエージェントが DQL によって学習する場合 Independent Deep Q Learners(IDQL) という.

## 2 マルチエージェント強化学習による信号制御

実際の環境には, 膨大な台数の信号機が設置されているため, すべての信号機を単体の強化学習によって制御することは難しい, そこで, 本研究では, IDQL を信号機制御に適用する. つまり, 各エージェントが各信号機を制御する.

### 2.1 強化学習の設定

信号機を強化学習によって制御するために，行動空間, 状態空間, 報酬について定義する.

**行動空間:** 南北の青の比率を $NS$, 東西の青の比率を $EW$ として，次の 3 つのスプリットを行動空間とする.

1. $NS = EW$
2. $NS > EW$
3. $NS < EW$

各エージェントはサイクル毎に, この 3 つの中から 1 つを選択する.

**状態空間:** 状態は, 各エージェントが制御している信号機の 1 サイクルの交差点の流入方向のレーン (流入レーン) の平均車両占有率と平均停車数により表現する.

**報酬:** 各エージェントが制御している信号機がある交差点の流入レーンの 1 サイクル $\Delta t$ の平均車両待ち時間を元に，以下のようにエージェント $i$ の報酬 $r^i$ を定義する,

---

**Toward Controlling the Traffic Lights using Multi-agent Reinforcement Learning**

Takuay OKANO†‡, Masaki ONISHI†‡and Itsuki Noda‡
†Department of Policy and Planning Sciences University of Tsukuba
Tsukuba, Ibaraki 305-0577, Japan
‡Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki 305-8560, Japan
okano565656@gmail.com

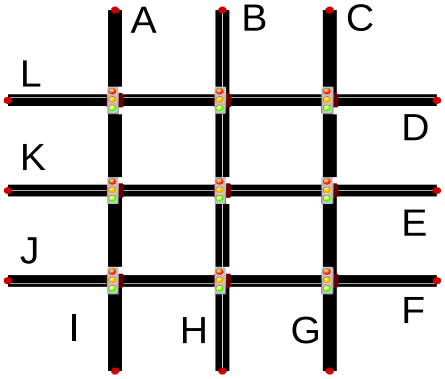


図 1: 実験で用いるマップ. アルファベットは車両の出発地点もしくは到着地点を表す

$$r^i = -W^i_{t,t+\Delta t}\,, \tag{1}$$

$$W^i_{t,t+\Delta t} = \frac{1}{\Delta t}\sum_{k=t}^{t+\Delta t}\frac{1}{|L^i|}\sum_{l\in L^i} w^l_k\,, \tag{2}$$

$L^i$ は信号 $i$ が制御している流入レーンの集合, $w^l_t$ は $t$ ステップ時のレーン $l \in L^i$ を走行している車両の待ち時間を表している. 各エージェントがコントロールしているレーン上にいる車両の待ち時間が少なければ少ないほど, 高い報酬を与える.

## 3 実験

先に述べた, マルチエージェント強化学習による信号機制御を交通シミュレーター SUMO(Simulation of Urban MObility) を用いて評価する.

### 3.1 実験設定

本研究では 3×3 のシンプルな格子状のマップを用いる. 実験で用いるマップを図 1に示す. 各レーンの長さは 100m, 制限速度は 40km/h, 1 サイクルを 90 秒とする. 表 1に本実験で用いる OD 表を示す. 比較手法として, $NS = ES$ を選択し続ける (Normal) と, ランダムで選択する (Random) を用いる. 1 エピソードをすべての車両が生成されてから, 目的地に到達するまでとし, 複数回同じ設定で実験する.

### 3.2 実験結果

実験結果を図 2に示す. 図 2から, IDQL によって信号機を制御することで, Normal, Random 制御に比べて平均待ち時間が減少していることがわかる. 序盤のエピソードでは学習が進んでいないため, 性能が向上していないが, エピソードが進むに連れて待ち時間が減少していることがわかる.

表 1: 実験で用いる OD 表

| O \ D | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 5 | 1 | 4 | 3 | 4 | 3 | 5 | 200 | 1 | 3 | 5 |
| B | 4 | 0 | 1 | 3 | 4 | 5 | 2 | 100 | 2 | 1 | 3 | 5 |
| C | 2 | 2 | 0 | 5 | 1 | 1 | 50 | 2 | 4 | 4 | 5 | 4 |
| D | 1 | 3 | 3 | 0 | 3 | 5 | 4 | 3 | 2 | 5 | 1 | 200 |
| E | 3 | 1 | 1 | 2 | 0 | 2 | 3 | 4 | 1 | 4 | 100 | 5 |
| F | 5 | 4 | 3 | 5 | 1 | 0 | 3 | 5 | 4 | 50 | 5 | 2 |
| G | 1 | 1 | 50 | 5 | 4 | 2 | 0 | 2 | 4 | 3 | 2 | 4 |
| H | 4 | 100 | 2 | 5 | 3 | 4 | 5 | 0 | 1 | 2 | 1 | 5 |
| I | 200 | 3 | 4 | 2 | 1 | 2 | 4 | 3 | 0 | 2 | 2 | 5 |
| J | 1 | 5 | 3 | 4 | 1 | 50 | 2 | 2 | 3 | 0 | 3 | 2 |
| K | 3 | 3 | 1 | 5 | 100 | 5 | 2 | 1 | 5 | 5 | 0 | 5 |
| L | 4 | 1 | 4 | 200 | 5 | 1 | 3 | 4 | 5 | 5 | 4 | 0 |

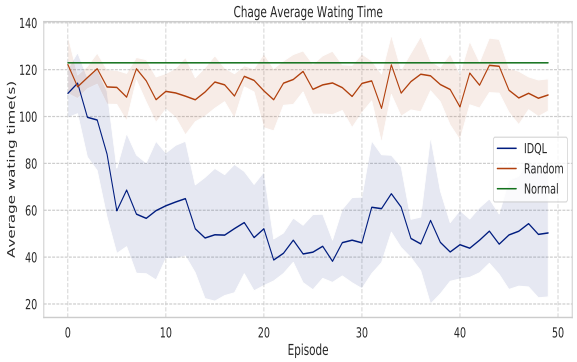


図 2: エピソード毎の平均待ち時間の推移. 薄い部分は標準偏差

## 4 まとめ

本稿では, 複数の信号機をマルチエージェント強化学習によって制御するための構成（行動空間, 状態空間, 報酬）を提示した. そして, シンプルなマップを用いて評価した. 結果として, 本設定においては, IDQL によって走行車両の待ち時間を減少するように信号機を制御できることを確認した.

しかし, 今回の実験で用いた設定は, 各エージェントが協調せずに, 性能を向上させることが可能な容易なものであった. 今後は, エージェント間での協調が必要な問題で実験を行う必要がある.

## 参考文献

[1] V. Mnih *et al.*, Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533.

[2] L. Busoniu, R. Babuska, B. De Schutter, A comprehensive survey of multiagent reinforcement learning, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38 (2) (2008) 156–172.