

## Deep Q-Network を用いた交通信号制御システムの提案

## Development of Traffic Control System by use of Deep Q-Network

佐藤季久恵 \*<sup>1</sup> 高屋英知 \*<sup>1</sup> 小川亮 \*<sup>1</sup> 芦原佑太 \*<sup>1</sup>\*<sup>2</sup> 栗原聡 \*<sup>1</sup>  
 Kikue Sato Eichi Takaya Ryo Pgawa Yuta Ashihara Satoshi Kurihara

\*<sup>1</sup>電気通信大学

The University of Electro-Communications

\*<sup>2</sup>株式会社クロスコンパス・インテリジェンス

XCompass Ltd.

Traffic congestions occurring in urban areas cause not only time losses for drivers but also economic losses. The basic method to solve this problem is the traffic signal control. It is challenging to improve the traffic signal control flexibility to adapt the real-time traffic change. Although some adaptive methods using genetic algorithm, multi-agent system, and so on, have been proposed in previous studies, those methods require hand-crafted information from traffic situation. In this study, we propose a traffic signal control technique using deep Q-network, which can automatically extract features from raw images of traffic situation and effectively find optimal traffic light control policies. We perform experiments with the traffic microsimulator SUMO and verify the effectiveness of the proposed algorithm.

## 1. はじめに

都市部で発生する交通渋滞は、ドライバーの時間的損失だけでなく、輸送の遅延や燃料消費の増加に伴う経済的損失を引き起こしている。更に大気汚染や追突事故などの主要な原因としても指摘されている [1]。交通量が所与のもとで交通渋滞を防ぐ手立てとしては、適切なナビゲーションによって交通量の分散を図るアプローチや、信号制御によりスムーズな交通流を生み出すアプローチなどがある。本研究では、後者の信号制御に焦点を当てる。

一般道路での信号制御は、交通量から算出したパラメータのパターンを複数用意し、時間帯によってそれらを使い分ける静的な制御方法である。そのため、交通事故やイベントの開催など、突発的な交通流の変化に対応できない。したがって、時々刻々と変化する交通流の特徴を的確に抽出し、それに対する最適なパラメータ操作を行うことが重要である。

そこで、高い特徴抽出能力を持つ Deep Learning と、報酬に基づいた最適な行動を学習する強化学習を組み合わせた Deep Q-Network(DQN) に注目する。特にビデオゲームや囲碁において、DQN は人間のプロさへも上回るほどの成果を上げている [2]。本研究では、この DQN を用いて信号機を動的に制御するシステムを提案する。

## 2. 関連研究

信号制御に関する研究は様々な角度から行われている。高橋ら [3] は、各交差点に存在する信号機をエージェントとし、エージェント同士が協調することで自律的に信号制御を行う手法を提案した。西原ら [4] は、「青、黄、赤」と表示される信号機をマルチエレメント GA を用いて適切に変化させ、効率的に信号制御を行う手法を提案した。Liu ら [5] は、隣接交差点の交通情報から Q 学習により信号機のパラメータを制御し、効率的に交通量を改善することができることを示した。Van ら [6] は、Deep Q-Network を信号制御に適用し、入力として車を 2 進数で表した行列を用いて信号制御を行う手法を提案し

連絡先: 佐藤季久恵, 電気通信大学, 東京都調布市調布ヶ丘 1-5-1 東 2 号館 417 栗原研究室, 042-443-5664, ksato@ni.is.uec.ac.jp

た。

いずれの先行研究も、高次元な交通情報からあらかじめ信号制御を用いるために必要な情報を抽出し、システムの入力値として与えている。しかし、抽出されていない情報にも信号制御を行う際に有効な情報が含まれている可能性がある。そこで、本研究では、画像という高次元な情報からエージェント自身が信号制御に必要な特徴を抽出し、適切なパラメータ操作を出力することを目的とした、DQN を用いた交通信号制御システムを提案する。

## 3. Deep Q-Network のアルゴリズム

DQN は、深層学習を用いた特徴抽出と強化学習における代表的な手法である Q 学習を組み合わせた技術である。Q 学習とは、状態  $s$  (入力) を与えたときに Q 値 (行動の価値) が最大となる行動  $a$ , すなわち、最も報酬  $r$  が得られると期待される行動 (出力) をとるようにエージェントが学習する方法である。本研究では入力値として画像を用いるが、画像情報をもとに Q 学習を行うためには、エージェントが過去の画像から得られる高次元な情報から特徴量を抽出し、活用しなければならない。そこで、DQN では画像からの特徴抽出に、深層学習の手法である Convolutional Neural Network(CNN) を用いている。また、下記の 2 つの手法によって、連続値に対しても Q 学習を適用できるような工夫がなされている。

## 1. experience replay

状態が変化するごとに  $(s_t, a_t, r_t, s_{t+1})$  を Memory 表に保存する。Memory 表からランダムに  $(s_t, a_t, r_t, s_{t+1})$  を選び出し、選び出した  $(s_t, a_t, r_t, s_{t+1})$  を教師データとして Q 値を更新する。Memory 表は、 $(s_t, a_t, r_t, s_{t+1})$  を 10,000 組保存できる構造となっており、随時更新していく。

## 2. neural fitted Q iteration

experience replay で更新された Q 値を毎回反映してしまうと収束が安定しない。そこで、10,000step ごとに Q 値の更新を行う。

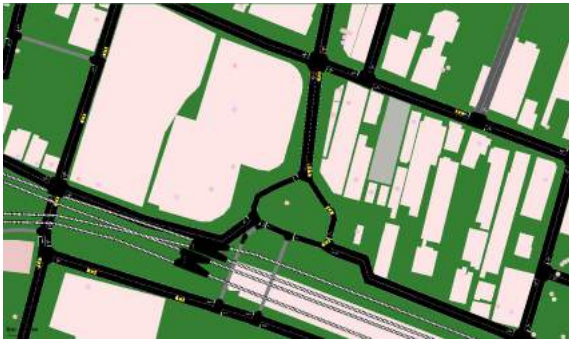


図 1: 調布駅周辺を示したシミュレーション画像

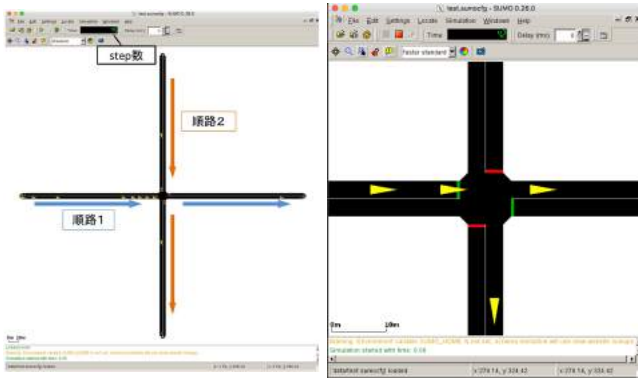


図 2: シミュレーションの全体像

図 3: 入力画像

## 4. 提案システム

本研究では、環境として交通シミュレータを、エージェントとして DQN を用いる。まず、環境とエージェント、それぞれの設定について述べる。

### 4.1 環境設定

#### 4.1.1 ミクロ交通流シミュレータ SUMO(Simulation of Urban MObility)

本研究では、ミクロ交通流シミュレータ SUMO を用いて実験を行う。SUMO はドイツ航空宇宙センターを中心に開発されているオープンソースのシミュレータである。車の速度、加速度、走行ルートや信号機や道路の設置など、自由に設定することが可能である。また、図 1 のように、実際の道路環境を表現することも可能である。

#### 4.1.2 道路ネットワークおよび入力画像

車は 3step に 1 台の頻度で出現し、左端から右端(順路 1)と、上端から下端(順路 2)までの距離をそれぞれ 300m とし、車の通過率を順路 1:順路 2 = 1:5 とした環境を設定する。なお、順路 1, 順路 2 はいずれも直進のみ行うものとする(図 2 参照)。本研究では、図 2 を拡大させた、図 3 を入力画像とし、この画像をもとに学習を行う。

### 4.2 エージェント(DQN)設定

次に、エージェント設定について述べる。1step ごとに SUMO の画像を撮り、その画像を  $87 \times 87$ (pixel) にリサイズし、グレースケールに変換したものを入力画像として与える。その画像を 4 枚集めて 1step ごとにエージェントに与え、それら

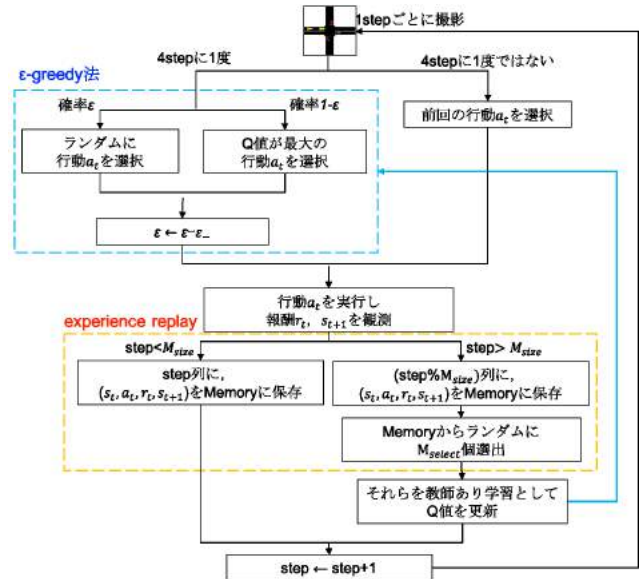


図 4: エージェント(DQN)のアルゴリズム

の画像を元に  $\epsilon$ -greedy 法<sup>\*1</sup>により行動  $a_t$  を選択し、実行する。その後、 $r_t, s_{t+1}$  の値を得、Memory 表に保存する。その後、Memory 表の中からランダムに選択し<sup>\*2</sup>、それらのデータをもとに Q 値を更新する。その後、次の step に移る(図 4 参照)。

## 5. 実験結果とその評価

車の待ち台数(waiting count)や待ち時間(waiting time)を負の報酬とし、報酬を与えるタイミングを遅らせるパラメータ Delay の値を変えて実験を行った(図 5, 6 参照)。waiting count を報酬とした実験では 2100 エピソード<sup>\*3</sup>、waiting time を報酬とした実験では 1500 エピソードのシミュレーションを行った。グラフは、Delay の値を 0,1,2 で比較した結果である。

### 5.1 実験結果

図 5, 6 をみると、400 エピソードを境に負の報酬が減少しているの、学習がうまくいっていることが分かる。また、グラフから、Delay=1 のとき最も精度がよいといえる。

### 5.2 静的な信号制御との比較

次に、DQN を用いた信号制御の精度を評価するために、車の交通流に合わせた静的な環境を用意し、DQN による手法で最も精度のよかった下記の 2 つの提案手法と比較手法(静的な信号制御)をそれぞれ比較する。

- 提案手法 1  
waiting count を負の報酬とし、Delay を 1 に設定。
- 提案手法 2  
waiting time を負の報酬とし、Delay を 1 に設定。
- 比較手法(静的な信号制御)  
比較手法として一定の間隔で信号機が切り替わる静的な

\*1  $\epsilon$  の確率で無作為に選択し  $1-\epsilon$  の確率で Q 値が最大の行動を選択

\*2 ただし、本実験では 32 個選択し、ミニバッチ学習を行った。

\*3 1 エピソードは 200step に等しい

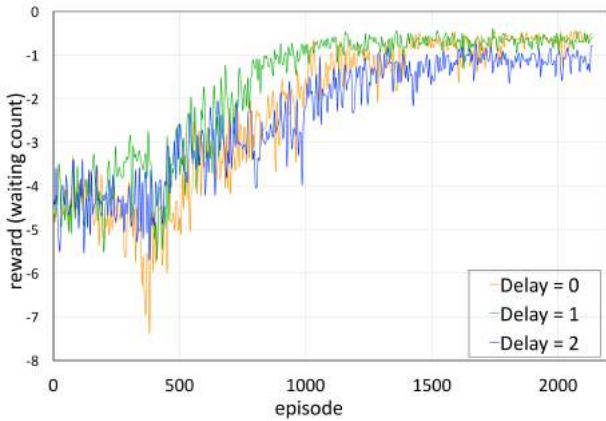


図 5: reward(waiting count)

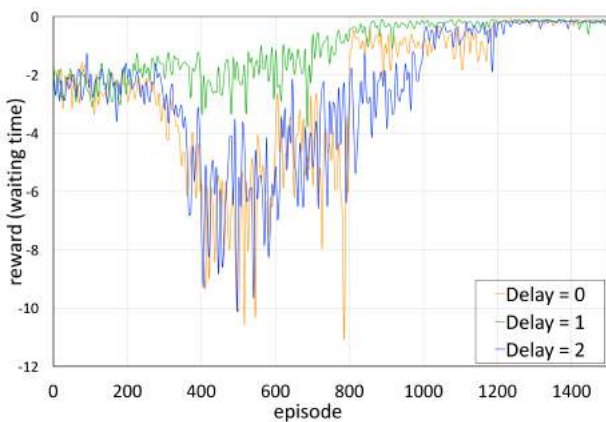


図 6: reward (waiting time)

信号制御を用いる。そのため、車の流量、黄色信号の時間等の道路環境は、提案手法と同じとし、信号機のサイクル<sup>\*4</sup>は50stepとする。また、信号機のスプリット<sup>\*5</sup>は車の通過率に合わせて、順路1を17%、順路2を83%とする。これらの実験設定の詳細を表1に示す。

### 5.3 提案手法 1,2 の評価

提案手法1では学習が収束している1300エピソード以降のデータを抽出し、提案手法2では850エピソード以降のデータを抽出し、そのデータをもとに評価を行う。提案手法1,2と比較手法をそれぞれ比較した結果を図7,8に示す。横軸はエピソード数、縦軸は1エピソードごとの負の報酬 (waiting count, waiting time) を蓄積させたものである。いずれも比較手法よりもDQNによる手法の方が、渋滞を解消できているといえる。

## 6. おわりに

本研究では、Deep Q-Networkを用いた交通信号制御システムを提案した。シミュレーションによる検証の結果、交通流の変化に対応した効率的な信号制御が行えることが確認された。

課題として、本研究では、サイクル長において現実を反映さ

\*4 信号機が赤→黄→青へと1巡する時間をサイクルと呼ぶ。

\*5 各信号で通行権を割り当てられる時間配分をスプリットと呼ぶ。

表 1: 静的な信号制御の環境設定

車の頻度	1台 / 3 step
車の通過率	順路1 : 順路2 = 1 : 5
順路1, 順路2の長さ	300 m
黄色信号の時間	2 step
1エピソードあたりのstep数	200 step
スプリット	順路1=17%, 順路2=83%
1エピソード間のcycle数	10 cycle
grgrのstep数	2.65 step
yryrのstep数	2 step
rgrgのstep数	13.35 step
ryryのstep数	2 step

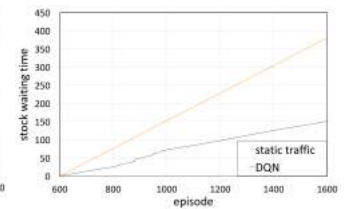
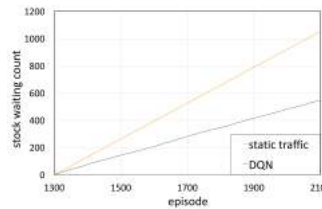


図 7: 提案手法 1 と比較手法 図 8: 提案手法 2 と比較手法

せた設定になっていないことや、ごく小規模な交差点での学習にとどまっていることがあげられる。よって、今後は、現実的な設定での学習や、大規模道路ネットワークを対象とした実験を行う予定である。

## 参考文献

- [1] 国土交通省生産性改革プロジェクト, 国土交通省生産性革命プロジェクト第1弾, 2016.
- [2] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013).
- [3] 高橋光紀, et al. "マルチエージェントモデルによる信号制御シミュレーションの考察." 研究報告知能システム (ICS) 2014.7 (2014): 1-6.
- [4] 西原稔貴, et al. "マルチエレメント GA による道路交通信号パラメータの最適化と実環境における検証 (ITS 画像処理, 映像メディア, 視覚及び一般)." 電子情報通信学会技術研究報告. IE, 画像工学 111.442 (2012): 263-268.
- [5] Liu, Weirong, et al. "Cooperative multi-agent traffic signal control system using fast gradient-descent function approximation for V2I networks." Communications (ICC), 2014 IEEE International Conference on. IEEE, 2014.
- [6] van der Pol, Elise. Deep reinforcement learning for coordination in traffic light control. Diss. Master's Thesis. University of Amsterdam, 2016.