

Deep Q-Network を用いた反社会的エージェントの 道路交通における行動学習シミュレーション機構の試作

日車 徹[†] 福田 直樹[‡]

[†] 静岡大学情報学部

[‡] 静岡大学大学院情報学領域

1 はじめに

本研究では、反社会的なエージェントが反社会的な行動を学習していくことでどのような悪影響が生じるか明らかにすることを目的とする。具体的な問題設定として、反社会的車と一般車が混在する道路交通において反社会的車が一般車の移動を妨害する行動を学習することを想定する。本稿ではその行動学習のシミュレータの試作について述べる。

シミュレータは道路の状況を任意に設定することができ、シミュレーションの進行の様子をリアルタイムかつグラフィカルに観測することができる。このシステムを用いることで反社会的車の行動学習を実験し、その様子を評価することができる。

本シミュレータを実現するために、行動学習の手法として、予め環境のルールを教えずに学習を進めることができる Deep Q-Network (DQN)[1] を導入する。

2 道路交通における反社会的行動の課題

自身の効用を向上させるために個人または全体の効用を低下させる人がある。例えば当たり屋のように意図的に他者と事故を起こし、相手を恐喝して賠償金を請求するという行為が挙げられる。

当たり屋行為により自分は利益を得ることができ、相手は時間的・金銭的・精神的に損害を受けるという問題が起こりうる。それに加えて事故の発生によってその道路の交通は滞り [2] 周囲の利用者も損害を受け、事故処理のために警察や消防、民間企業にも追加のコストを強いることとなりうる。

A Preliminary Approach on Learning Anti-social Behaviors in road traffic on a Simulator using a Deep Q-Network

Toru HIGURUMA[†] and Naoki FUKUTA[‡]

[†] Faculty of Information, Shizuoka University
432-8011, Hamamatsu, Japan

[‡] College of Informatics, Academic Institute,
Shizuoka University

432-8011, Hamamatsu, Japan

[†] cs14084@cs.inf.shizuoka.ac.jp

[‡] fukuta@cs.inf.shizuoka.ac.jp

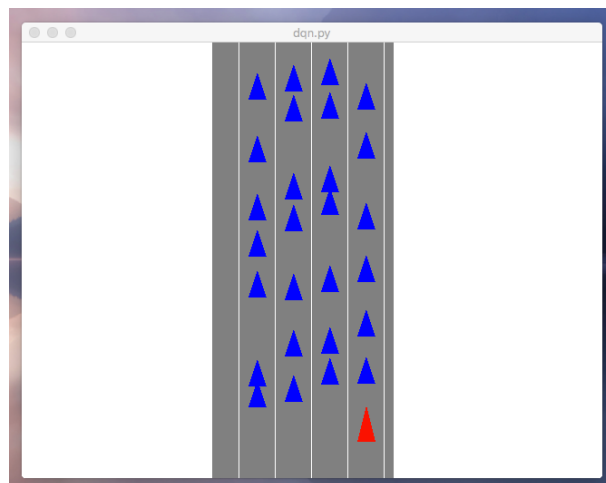


図 1: 試作中のシミュレータの画面例

平成 29 年警察白書¹によると、当たり屋による平成 28 年の検挙数は 79 件、被害額は 1,121 万円となっている。この数値はあくまで検挙に至ったもののみの集計であり、検挙に至らなかった被害を含めるとさらに大きな数値となることが推察できる。

本研究では以上のような社会的規範を逸脱した行動を反社会的行動と定義し、議論の対象とする。

3 シミュレーションの問題設定

実装するシミュレータの画面を図 1 に示す。シミュレータ上にはユーザから与えられた道路が再現される。道路上には反社会的車（図 1 中赤色の三角形）が 1 台と一般車（図 1 中青色の三角形）が複数台配置される。このシミュレータは道路上で反社会的車と一般車が自律的に移動するという状況を再現する。

道路上の車は反社会的車と一般車の 2 つに大別され、それぞれ特定のアルゴリズムに従って行動選択を行う。一般車はあらかじめ作成したアルゴリズムに従って行動する。

¹ <https://www.npa.go.jp/hakusyo/h29/toukei/toku/27.xlsx>

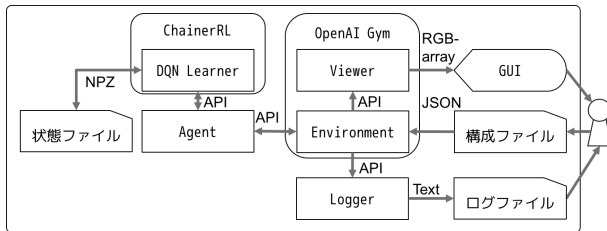


図 2: シミュレータの構成図

一方、反社会的車は Deep Q-Network (DQN) による行動学習を行う。反社会的車は環境上で様々な行動を試行し、その報酬として環境から一般車の効用を下げた合計値を得る。行動から得た報酬から次の行動を修正していくことで反社会的車は行動学習を進めていく。

4 シミュレータの構成

シミュレータの構成を図 2 に示す。シミュレータはシミュレートする環境そのものを表す Environment を中心に、Agent, Deep Q-Network, Viewer, Logger から構成される。

Environment はシミュレーション対象の道路そのものおよび道路上の一般車の動作を管理する。またシミュレーションそのものの進行に必要な API を実装し、サブシステムに公開する。Agent からは反社会的車の行動を受け取り、それに応じてシミュレーションを進行し、結果から報酬を計算して Agent に返す。シミュレーションの結果は Viewer および Logger にも引き渡し、Environment 自身の情報を可視化できるようにする。

Viewer は Environment 上の道路と各車の情報を受け取り、それに応じて GUI 上に表示可能な形式として道路状況を可視化する。可視化した画像は RGB 配列として GUI 表示以外でも利用することができる。このため可視化した情報を内外のプログラムとやりとりすることが可能である。

Logger は Environment 上のシミュレーションの進行状況を受け取り、そのログをテキスト形式で逐次出力する。Viewer の画面出力と異なり、内部の数値そのものを各タイムステップごとに記録することができる。このためシミュレーションの進行する様子を後から解析するために使用することが可能である。

DQN Learner は Deep Q-Network を用いた学習器で

ある。Agent がシミュレータ上で取る行動はこの DQN Learner が決定し、Environment に受け渡される。その行動に対する報酬は Environment から Agent を通して DQN Learner に返される。行動選択とその報酬の評価を繰り返し、DQN Learner は学習を進めていく。

Agent は Environment 上の反社会的車の動作を管理する。実際には行動選択を DQN Learner に委譲し、Environment と DQN Learner 間の通信を管理する。また本シミュレータのエントリーポイントとしても機能する。

このシステムのユーザはシステムへの入力として、シミュレートする環境についての構成ファイルを用意する必要がある。構成ファイルには道路の車線数と各車線の属性、一般車の属性および初期位置が含まれる。この構成ファイルは JSON (JavaScript Object Notation) 形式²で表現される。このためユーザは任意のエディタでシミュレートする環境を構成することができる。Environment はこのファイルを読み込み、シミュレートする環境を再現する。

またユーザはシステムからの出力として Logger から逐次出力されるログファイルおよび Viewer から逐次出力される GUI イメージを得る。これらを基にユーザは反社会的車の動作を観測し、その行動学習を評価することができる。

5 結論と今後の課題

本稿では、エージェントが反社会的な行動を学習するシミュレーション機構の試作について述べた。今後の課題としては、他の強化学習手法を導入することでシミュレータの評価及び分析を行うことである。

参考文献

- [1] Volodymyr Mnih, Koray Kavukcuoglu, et al. (2013) “Playing Atari with Deep Reinforcement Learning,” NIPS Deep Learning Workshop.
- [2] Guni Sharon, Josiah P. Hanna, et al. (2017) “Real-time Adaptive Tolling Scheme for Optimized Social Welfare in Traffic Networks,” Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8-12, 2017, São Paulo, Brazil, pp.828-836.

²<https://json.org>