

被災地の復旧フェーズを考慮した交通流シミュレーションに基づく深層強化学習による道路修復計画策定

今井星太¹⁾, 宮村倫司²⁾

Shota Imai and Tomoshi Miyamura

1) 日本大学大学院工学研究科 大学院生 (〒963-8642 福島県郡山市田村町徳定字中河原 1, E-mail: cesh22004@g.nihon-u.ac.jp)

2) 工博 日本大学工学部 准教授 (〒963-8642 福島県郡山市田村町徳定字中河原 1, E-mail: miyamura.tomoshi@nihon-u.ac.jp)

During a disaster, routes to destinations may be blocked due to road damage, making repairs necessary. Emergency activities in a disaster-stricken area involve three phases: saving lives, transporting emergency supplies, and restoring traffic for general vehicles. In this study, a traffic flow simulator was used to simulate traffic flows during each phase, and vehicle arrival times were evaluated. Based on the simulations, deep reinforcement learning was used to determine where road repairs are necessary for each phase, and a repair plan was designed accordingly.

Key Words : Traffic Flow Simulation, Deep Reinforcement Learning, Disaster Recovery

1. はじめに

地震などの被災地で道路が破損した際には、被災地へ出入りするルートを確認することが困難になるため、修復が必要となる。災害時の被災地での応急活動には 3 つのフェーズがあり、人命救助、緊急物資輸送、一般車両通行の順に進められる^[1]。各フェーズにおいて、破損した道路をどのような順番で修復するのかについての計画を立案する必要がある。

交通流には様々な要因が複合的に関連している。交通現象そのものが異なる属性や目的をもつ多数の車両の相互作用の結果である。この問題に対して吉村ら^[2]、藤井ら^[3]はマルチエージェントシミュレーションによるアプローチを提案した。また、藤井らは知的マルチエージェント交通流シミュレータ ADVENTURE_Mates^[4]を開発している。

強化学習とは教師あり学習、教師なし学習と並ぶ機械学習の分類の一つである。強化学習では、タスクを実行する上で、単体の行動結果ではなくタスク全体の行動結果を評価することにより、一連の意思決定を学習する。深層強化学習は強化学習の一種であり、強化学習の行動結果の評価にニューラルネットワークを適用した学習手法である^[5]。近年、深層強化学習を用いた行動計画作成の研究が行われている^[6]。荻野は強化学習を粒子の充填問題に応用した^[7]。また、小寺等は強化学習をトラス構造物の施工経路生成に応用した^[8]。

本研究では藤井らが開発した交通流シミュレータを用いて災害時の交通流をシミュレートし、車両の到達時間を評価する。それに基づいた深層強化学習により破損した道路の修復の順番を決めることで、修復計画の作成を行う。

2. 被災地の応急活動フェーズ

災害時の被災地での応急活動には 3 つのフェーズがある。フェーズは図 1 のように遷移していく。災害後 72 時間後程度までは、人命救助フェーズと呼ばれる。傷病者が多数発生し、早期の救出救助活動が開始され、救助された多数の傷病者を医療機関に搬送される。また、ライフラインや交通機関が途絶し、被災地外からの支援の受け入れが少ない状態である。72 時間~1 週間後程度までは緊急物資輸送フェーズと呼ばれる。このフェーズでは、ライフライン等が復旧し始め、人的・物的支援の受入体制が確立されている状態である。これ以降は一般車両通行フェーズと呼ばれ、地域医療やライフライン機能、交通機関等が徐々に復旧している状態である。

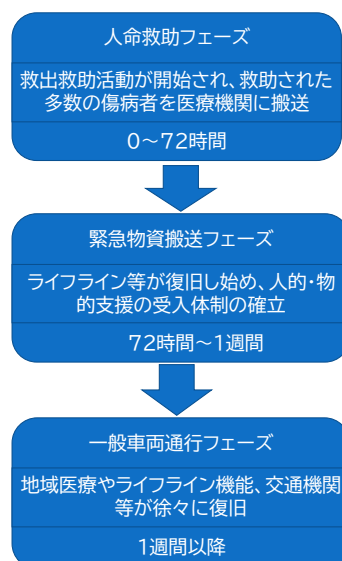


図-1 応急活動フェーズ

3. 交通流シミュレータ

(1) ADVENTURE_MATES

ADVENTURE_MatesはADVENTUREプロジェクト^[9]において開発されたモジュールのひとつであり、知的マルチエージェントモデルに基づく交通流シミュレータである。知的エージェントとしてモデル化された車両は道路環境中で他の車両エージェントと相互作用しながら自律的に意思決定(経路探索, 加速度決定)をする。複雑な交通現象は知的エージェント同士の相互作用の結果として創発するよう設計されている。

(2) 設定可能な情報

シミュレーションを行う際、信号や交通量を設定する必要がある。ADVENTURE_Matesでは、信号ごとに信号パターンを決定できる。また、各ネットワークの端点が流入流出点となっているため、そこからの車両発生台数を設定することで交通量を制御できる。

(3) osm2mates

OSMとは、OpenStreetMap^[10]の略称である。OpenStreetMapとは、自由に利用でき、なおかつ編集機能のある世界地図を作る共同作業プロジェクトである。osm2matesはADVENTURE_Matesを実行するための地図データ(mapPosition.txt, network.txt)をOpenStreetMapから自動的に作成するプログラムである。

4. 強化学習

強化学習は行動心理学者であるスキナーが動物の行動を説明する基本原理として唱えた強化に由来する考え方である^[11]。代表的な例であるスキナー箱を用いて、基本的な考え方を説明する。図-2のネズミは最初、無造作な行動をする。しかし、ある時偶然電源スイッチを押し、レバーを押したことで餌(報酬)を得ることができる。このような動作を繰り返すうちにネズミはスイッチの状態によって、餌(報酬)を得るために適切なレバーやスイッチを押すこと(行動)を覚える。強化学習はこれを確率的動的計画法の枠組みで定式化したものである。

強化学習では環境という行動と行動による状態の変化が定義されており、ある状態への到達に対し報酬が与えられる空間がある。この環境が開始されてから終了されるまでを1エピソードといい、この1エピソードの中で得られる報酬を最大化することが強化学習の目的である。

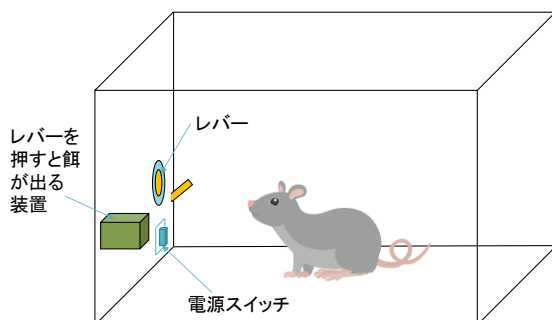


図-2 スキナー箱

(1) Q学習

状態による行動価値(価値関数)をQ値と呼ぶ。Q値はQテーブルと呼ばれる表に記録をし、行動をするたびに該当するQ値を更新する。状態 s_t 、行動 a 、報酬 r 、学習率 α 、割引率 γ としたとき、Qテーブルの更新にはベルマン方程式に基づく式(1)を用いる。

$$Q(s_t, a) = (1 - \alpha)Q(s_t, a) + \alpha(r + \gamma Q(s_{t+1}, a')) \quad (1)$$

(2) 深層強化学習(Deep Q Network (DQN))

Deep Q Network (DQN)とは、Q学習のQ値をQテーブルではなく、ニューラルネットワークで計算する学習方法である。ニューラルネットワークとは、人間の脳を模した数理モデルである。ニューラルネットワークは入力層と一つ以上の中間層、出力層を含む、複数のノード層で構成されている。各ノードは別のノードに接続し、重みと閾値を持つ。深層強化学習では、ニューラルネットワークの入力を状態とし、出力を各行動のQ値とする。Q学習におけるベルマン方程式を教師データとすると、損失関数を二乗誤差により定める。これを最急降下法を基にしたアルゴリズムにより最小化することでQ値を算出する関数として使えるニューラルネットワークを生成する。

5. 深層強化学習による道路修復計画の作成

(1) 深層強化学習による災害時道路修復計画の作成手法の概要

修復途中の道路の評価はADVENTURE_Matesを用いて行う。いくつかの地点で発生させた車両が目的地に到達するのにかかる時間を基に報酬に計算する。一方、道路の修復費用を負の報酬として評価する。これらの報酬の和が大きくなるように、図-3のような深層強化学習によって道路の修復の順番を決定することで道路修復計画を作成する。

シミュレーション上ではフェーズの表現を車両の種類と流入点の数によって表現する。車両は目的地が病院の緊急車両と目的地がランダムな一般車両の2つを用意する。人命救助フェーズでは、傷病者を搬送する緊急車両の交通流を再現する。緊急物資搬送フェーズでは、ボランティアや物資の受け入れを再現する。一般車両通行フェーズでは、平常な交通流を再現する。フェーズが進むにつれ、流入点の数とそれぞれの流入点から発生する車両の数を増やす。また、それぞれのフェーズで修復できる道路の数に上限を設ける。

破損した道の数を n 個としたとき n 個の各道路の状態は未修復と修復の2個であるため、全ての修復パターンの状態は 2^n 個となる。また、フェーズの状態は3個である。1エピソード内での行動の上限を m としたとき、修復を

行った回数の状態は m 個となる．そのため，状態の総数は $2^n \times 3 \times m$ となる．状態は $s_1 \sim s_{2^n \times 3 \times m}$ となる．行動はそれぞれの道の修復をするという行動と現在のフェーズの道の修復を終了するという行動とし， $a_1 \sim a_{n+1}$ とした．1 エピソードは 3 フェーズ目の道路の修復を終了したときか，行動の数が m を超えたときに終了するようにした．報酬は以下のように与えた．修復を行ったときに工事費を表す報酬を与えた．修復したときに現在のフェーズ中に修復できる道路の上限を超過していたら一定の負の報酬を与えた．フェーズを終了するという行動をとったときは車両全体に対して，（修復後の経済的損失－初期状態の経済的損失）の報酬を与える．経済的損失 L はフェーズの日数 d ，総走行時間 t ，人口 p ，平均時給 w ，1 時間当たりのガソリン代 c により式(2)で求める．

$$L = dtp(-w - c)$$

(2)

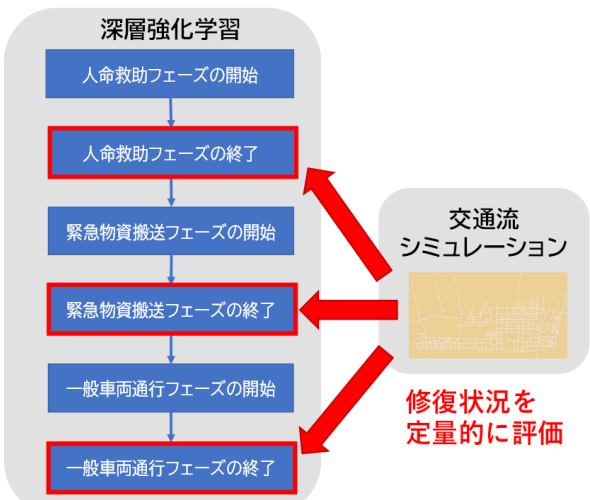


図-3 深層強化学習でのシミュレーション結果の利用

(2) DQN 用ニューラルネットワークの概要

ニューラルネットワークでは入力層のノード数を $n+2$ 個，中間層のノード数を 128 個で層は 3 層，出力層のノード数を $n+1$ 個とする．中間層の活性化関数には ReLU を使用する．Optimizer には adam を使用する．ニューラルネットワークへの入力， n 番目の入力ノードまでは対応する道路が修復されていれば 1，修復されてなければ 0 を入力する．また， $n+1$ 番目の入力ノードには現在のフェーズが何フェーズ目かを入力し， $n+2$ 番目の入力ノードには現在のフェーズで道路を修復した回数を入力する．

(3) 実装

本研究では，Python と Tensorflow を用いて実装を行う．深層強化学習での行動選択や報酬の計算，強化学習部分については Tensorflow の強化学習用ライブラリ TF-Agents を用いて実装する．Q 学習を行う深層学習部分について

は Tensorflow により実装を行う．交通流シミュレーションの入力データとなるマップの修復状況の変化に応じた書き換えや，交通流シミュレーションの実行といった深層強化学習以外の部分も Python により実装する．

6. 郡山駅周辺の道路修復計画作成への応用

(1) 問題設定と入力データの作成

郡山駅周辺が被災したと仮定し，その道路修復計画を提案手法により作成する．交通流シミュレーションの入力データとなるマップは，osm2mates により OpenStreetMap を基に作成した．郡山駅周辺から開成山公園周辺までを対象としている．ここでは，図-4 のように 11 本の道路が破損したと仮定して，それらの道路を削除した初期状態のマップを作成した．図-4 の中の赤の線が引かれている道が破損している道である．本研究ではすべての車線は片側 1 車線であると仮定している．また，信号に隣接している交差点の数ごとに，デフォルトの信号パターンを設定した．報酬の計算には郡山市のデータを基に作成した表-1 の値を使用する．修復箇所数の上限は人命救助フェーズでは 2，緊急物資搬送フェーズでは 4，一般車両通行フェーズでは 10 とし，1 エピソード内での行動の上限 m を 50 とする．また，人命救助フェーズでは図-5，緊急物資搬送フェーズでは図-6，一般車両通行フェーズでは図-7 のマップを使用する．車両の流入点を大きな青の点とし，救急車の目的地である病院は大きな緑の点で表している．

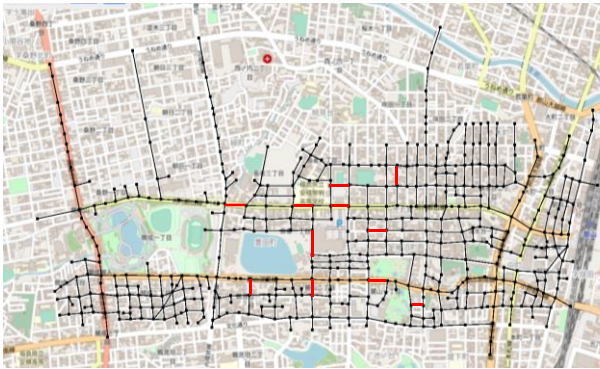


図-4 郡山駅前のマップ

表-1 報酬の計算に使用した値

項目	数値
道路の修復費用	60 万円
範囲内の人口	3 万人
平均時給	2440 円
1 時間当たりのガソリン代	495 円
人命救助フェーズの日数	3 日
緊急物資搬送フェーズの日数	4 日
一般車両通行フェーズの日数	20 日

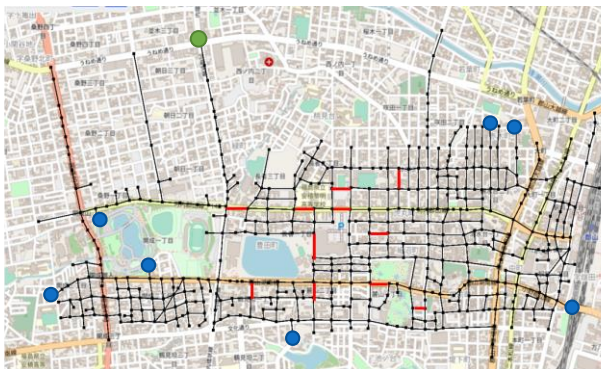


図-5 人命救助フェーズのマップ

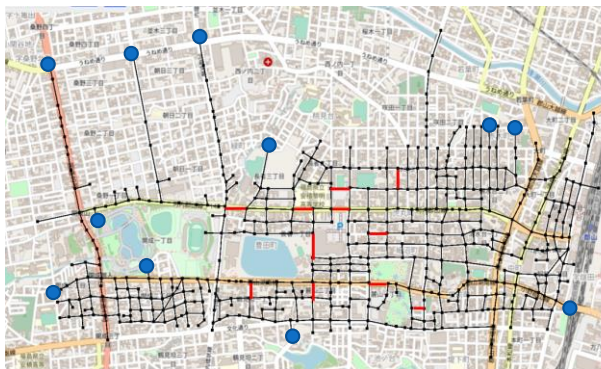


図-6 緊急物資搬送フェーズのマップ

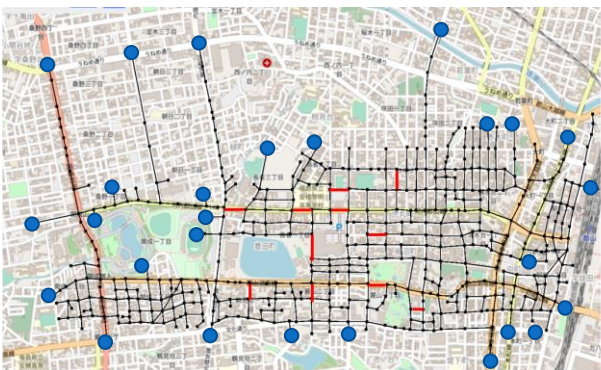


図-7 一般車両通行フェーズのマップ

(2) 深層強化学習による道路修復計画作成の結果

ここでは、予め修復過程の道路の全パターンのマップによるシミュレーションをシミュレーション内時間で1200s 分行って結果データを保存し、学習においてはそのデータを用いる。シミュレーションは3フェーズごとに 2^{11} パターンあるため、全6144パターンとなる。シミュレーションの実行時間は人命救助フェーズでは約35s、緊急物資搬送フェーズでは約40s、一般車両通行フェーズでは、170[s]となった。エピソード数2000、エポック数50で実行をしたところ、計算時間は5508.9s、2000エピソード目の報酬が122.65、使用されたシミュレーションは2093パターンとなった。各エピソードの報酬は、図-8のようになった。人命救助フェーズのシミュレーションが514パターン、緊急物資搬送フェーズのシミュレーションが861パターン、一般車両通行フェーズのシミュ

レーションが718パターン行われた。また、人命救助フェーズの修復箇所を図-9、緊急物資搬送フェーズの修復箇所を図-10、一般車両通行フェーズの修復箇所を図-11に示す。修復箇所を青い線、破損箇所が赤い線で表す。

学習により作成した修復計画では、修復されない道路があった。この原因は本研究で使用したマップには迂回路が多く存在するためだと考えられる。また、シミュレーションは、シミュレーション内時間で1200sを1回のみしか行っていないため、目的地がランダムな車両の走行時間を十分に評価できていないと考える。

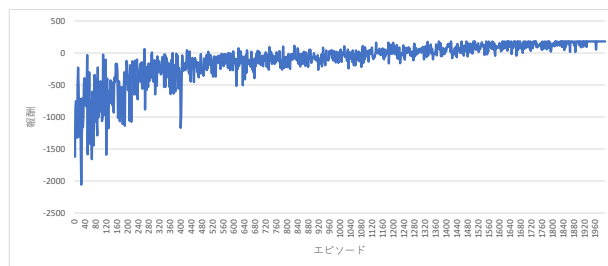


図-8 1エピソード毎の報酬



図-9 人命救助フェーズの修復箇所



図-10 緊急物資搬送通行フェーズのマップ



図-11 一般車両通行フェーズのマップ

7. おわりに

本研究では車両の到達時間の評価に基づいたDQNとQ学習により破損した道路のフェーズごとの修復箇所を決めることで修復計画の作成をした。報酬が高くなり易い道路は早いフェーズで修復されることが確認できた。修復計画の作成時間の大半をシミュレーションの計算が占めているため、その高速化が必要である。

また、フェーズごとの交通量や車両のパラメータに実際に計測したデータを使う必要があると考える。

謝辞: 本論文の作成にあたり、ご協力ご助言頂いた学部生には感謝いたします。

参考文献

- [1] 国土交通省: 災害時の通行可能な道路の確保と情報の取扱.
- [2] 吉村忍, 西川紘史, 守安智: 知的マルチエージェント交通流シミュレータMATESの開発, シミュレーション.23(3), 2004, pp.58-67.
- [3] 藤井 秀樹, 仲間 豊, 吉村 忍: 知的マルチエージェント交通流シミュレータMATESの開発 第二報: 歩行者エージェントの実装と歩車相互作用の理論・実測値との比較, シミュレーション, vol. 25, no. 4, pp. 274-280, 2006.
- [4] <https://adventure.sys.t.u-tokyo.ac.jp/download/Mates.html>
- [5] 牧野浩二, 西崎博光: TensorFlowによる深層強化学習入門 —OpenAI Gym+PyBullet によるシミュレーション—, 2021年, オーム社.
- [6] 松原崇充: 強化学習による行動学習—未知問題を閉じる道具—, 日本ロボット学会誌Vol.36, 2018.
- [7] 荻野正雄: 深層強化学習を用いた粒子ランダム稠密充填の検討, 計算工学講演会論文集, Vol. 26 (online), 2021.
- [8] 小寺 正也, 大崎 純, 林 和希: 強化学習とグラフ埋め込みによるトラスの安定な施工経路生成, 日本建築学会大会[東海], 2021年09月09日.
- [9] <https://adventure.sys.t.u-tokyo.ac.jp/jp/>
- [10] <https://openstreetmap.jp/>
- [11] 吉本 潤一郎, 銅谷 賢治, 石井 信: 強化学習の基礎理論と応用, 計測と制御, 第44巻, 第5号 2005年5月号.