

道路の防災機能評価に基づく 最適道路整備順序の決定アルゴリズム

柳沼 秀樹¹・陳野 由²・寺部 慎太郎³・田中 皓介⁴

¹正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)
E-mail: yaginuma@rs.noda.tus.ac.jp

²学生会員 東京工業大学大学院 環境・社会理工学院土木・環境工学系
(〒152-8552 東京都目黒区大岡山2-12-1)

³正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)
E-mail: terabe@rs.tus.ac.jp

⁴正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)
E-mail: tanaka.k@rs.tus.ac.jp

東日本大震災を契機として、道路ネットワークの耐災害信頼性を計測する手法の開発と運用が行われており、「広域拠点間の接続性」と「事業化区間の効率性」の観点から防災機能が評価されている。当該手法では、整備対象路線を与件として防災機能の評価を行っているが、整備すべき道路区間の抽出やそれらの整備順序を決定することができない。本研究では、防災機能評価を下敷きとして、深層強化学習を援用した最適な整備順序を導出するアルゴリズムを提案する。これにより、防災戦略上の優先改善候補ならびにそれらの整備順序を把握することが可能となり、地域の耐災害信頼性の向上に資することが期待される。

Key Words : *Disaster prevention planning, Road improvement, Deep reinforcement learning, Actor-critic*

1. はじめに

我が国では、近年、地震や豪雨などの自然災害が頻発しており、社会基盤施設がダメージを受けることによって、地域経済や市民生活に深刻な被害をもたらしている。道路や鉄道をはじめとする交通ネットワークは、発災初期における道路啓開や避難行動、その後の物資やボランティアの輸送など重要な役割を担っている。そのため、災害を想定したネットワークの整備や強化は喫緊の課題であり、災害による被害を最小限に抑えて、高いパフォーマンスを発揮する交通整備が求められる。

従前の道路整備において、時間短縮効果などを主な便益とする費用便益分析が評価の中心的な役割を担ってきた。近年では、様々な定量的・定性的な指標を導入したワイダーインパクトによる評価が試みられているが、基本的には平時における効果が計上されており、災害時における効果を明示的に扱うケースは少ない。そのような中で、2011年に発生した東日本大震災を契機として、我が国では道路ネットワークの耐災害信頼性を計測する防災機能評価手法が開発されており、災害時に道路が発揮

する効果を接続信頼性の観点から評価している。国土交通省道路局では、この手法を新規路線整備評価の実務に展開しており、費用便益分析と併用して政策決定の判断材料に利用されている。防災機能評価は「広域拠点間の接続性」と「事業化区間の効率性」の2つの観点から手法が構築されており、評価対象となる事業化区間の道路リンクについて、様々な災害シナリオでの耐災害信頼性を計測している。これにより、脆弱性の高い拠点ノードや改善効果の高い事業区間リンクを把握することが可能となる。

一方、財政的な制約から新規事業と並行して、既存ネットワークの耐災害性強化を進める必要がある。しかしながら、強化すべきノードやリンクの候補が多く存在するため、これらの中から効果の高い改良すべき候補を特定することは困難である。このような問題は、交通分野ではネットワークデザイン問題として研究が行われており、上位問題を供給サイドによるネットワークの改善、下位問題を需要サイドによる交通量配分とする二段階最適化によるアプローチが見られる。しかしながら、整備すべき道路リンクの候補集合を特定する既往研究は見ら

れるが、整備順序にまで言及した事例は見られない。

以上を踏まえて、本研究では道路ネットワークの耐災害信頼性に資することを目的として、改良を行うべき候補である災害時にクリティカルなリンクを抽出し、かつ最適なリンク整備順序を提案するモデルを構築する。方法論を提案する。具体的には、平常時における限られた年間予算内で早期に耐災害性の効果が発現できるリンク整備順序の決定する。整備効果を計る指標として、国土交通省が道路ネットワークの脆弱性を計測するために使用している「防災機能評価手法」を採用する。また、提案モデルのパフォーマンスを評価するために、全探索が可能な小規模のネットワークでの精度比較検証、ならびに Sioux Falls ネットワークに適用して計算可能性と手法有効性を検証する。本研究の成果として、防災戦略上の優先改善候補と整備順序を把握することが可能となり、地域の耐災害信頼性の向上に資することが期待される。

2. 既往研究の整理と本研究の位置づけ

(1) 道路の防災機能評価

防災機能評価は、2011年秋に「道路の防災機能評価手法（暫定案）」として運用を開始しており、東日本大震災の被災地での適用が最初の事例となる。その後、2015年に改良が加えられた「新たな道路の防災機能評価手法」として現在まで運用されており、災害時における拠点間の脆弱性および事業の効率性を評価するものである。

災害シナリオ（地震、豪雨、火山など）の下で、主要地方道以上を対象に災害有無ならびに整備有無の4パターンネットワークを構築する。なお、災害時のネットワークは、当該リンクが津波浸水区域や未耐震橋梁の有無などの基準を満たすか否かで被災有無を確定的に決定している。拠点は、被災拠点と支援拠点を地域の実情に応じて設定する。その上で、道路ネットワークの連結信頼性として各リンクの所要時間を用いた「脆弱度」および「改善度」を防災拠点ごとに定義している。

具体的には、災害時と平常時の各拠点間について多重性を考慮した複数経路の期待所要時間の比から脆弱度、新規プロジェクトを行う前後の各拠点間の期待所要時間の比から改善度を算出している。この値に閾値を適用することで、AからDの4段階でのランクづけを行い、ランクが低い（Dランク）拠点間に対して道路リンクの新規整備や修復の方針を検討している。しかしながら、どのリンクをどのタイミングで整備もしくは補修すれば最も効果的な整備ができるのかは決定することができない。そのため、防災機能評価にベースとして、整備もしくは補修すべきリンクの最適な選択手法を考える必要があるが、現時点では存在していない。

(2) 災害を想定したネットワークデザイン

本論に関連する既往研究として、災害時を想定した二段階最適化に基づくネットワークデザイン問題が挙げられる。これは、上位問題を供給サイドによるネットワークの改善、下位問題を需要サイドによる交通量配分とするアプローチである。代表的な既往研究として、長江ら⁵⁾は将来起こりうる大規模震災による道路ネットワークの被害に対して、リンク途絶による交通不便益を指標にクロスエントロピー法(CE法)を用いて最適耐震化戦略の手法を提案している。これは震災によるリンクの被災パターンを生起確率に基づいて複数個生成した上で、より不便益が大きい被災パターンの組み合わせを抽出することで、優先的な整備を行う箇所を特定している。また、柳沼らは長江らの手法を踏襲しながら、防災機能評価を導入した手法を構築している。

しかしながら、CE法で導出できる最適化戦略は修復すべきリンクの組み合わせ、即ち集合であり、整備順序を導出することは出来ない。

(3) 本研究の位置づけ

以上より、道路ネットワーク強靱化のためのリンクの耐災害化整備や新規に整備するリンクの強靱化効果に対する評価とそれに基づく優先度付は可能になっているものの、具体的なリンク整備の最適順序までは決定する方法が存在しない。そこで本研究では、各整備対象リンクを実際に順番に整備しながらフィードバックを得ることで各場面における最適行動を学習していく強化学習手法を採用する。また、各リンクの整備パターンはリンクの階乗であり、整備対象のリンクの数によっては膨大な状態数である。そのためニューラルネットワークによる関数近似値を用いた深層学習を組み込み、深層強化学習手法として最適整備順序決定モデルを作成する。これにより最終的に耐災害信頼性向上のための整備指針戦略を提案することを本研究の目的とする。

3. 最適道路リンク整備順序決定モデルの構築

(1) ネットワークの設定

はじめにネットワークの設定を示す。対象となる道路ネットワーク構造はノード集合 \mathcal{N} とリンク集合 \mathcal{L} からなる。このネットワーク上に災害時の一時避難所、待機場所として市役所や病院、道の駅などの防災拠点集合 \mathcal{H} が存在し、各防災拠点とノードを接続するダミーリンク集合 \mathcal{D} を定義する。また、対象ネットワークには、複数の災害シナリオが $s \in \mathcal{S}$ が存在しており、それぞれに対象ネットワークのリンク $l \in \mathcal{L}$ 上の交通施設集合 \mathcal{F}_l が災害により破壊される被災パターン（ \mathcal{F}_l 内の交通施設のいずれ

かが被災し通行不可の場合は1, 災害 s に影響を受けず破壊されない場合は0) が定められる. この時, 災害 s により被災する交通施設は, 本来は確率的であるが, 今回はいくつかの条件設定の有無による決定的を仮定する.

災害 s の被災パターンに指定されたリンクをハザードリンク $l_h \in \mathcal{L}_h (\mathcal{L}_h \in \mathcal{L})$ と設定し, 災害 s 発生時に被災しないリンク集合 $\mathcal{L}_r = \mathcal{L} - \mathcal{L}_h$ を残存リンク集合として定義する. ハザードリンクは, 耐災害化整備を行うことで改善され, 災害時においても被災しないとする. さらに, 対象ネットワークには, ミッシングリンク等によって未接続の新規事業化リンクが存在することがあるため, 新規整備対象リンク集合 \mathcal{L}_n を定義する. この新規整備対象リンクは, 整備後直ちに対象ネットワークのリンク集合 \mathcal{L} に追加され, 耐災害化整備を同時に行ったとみなしてハザードリンクに指定されないものと仮定する.

(2) 防災機能評価を援用した評価指標

対象ネットワークの脆弱性の評価は, 式(1)に示す脆弱度を用いるとし, 各防災拠点間の災害時と通常時の期待所要時間比より算出する. 期待所要時間は, 確率的な経路選択を適切に評価するために拠点間の多重性を考慮しており, 複数経路の所要時間をランダム効用理論に基づくログサム変数として下記のような式で算出する.

$$Z_{od} = 1 - \frac{W_{od}^u}{W_{od}^s} \quad (1)$$

$$W_{od}^u = -\frac{1}{\theta_{od}} \ln \sum_{k \in C_{od}^u} \exp\{-\theta_{od} \cdot t_{od,k}\} \quad (2)$$

$$W_{od}^s = -\frac{1}{\theta_{od}} \ln \sum_{k \in C_{od}^s} \exp\{-\theta_{od} \cdot t_{od,k}\} \quad (3)$$

ここで, Z_{od} は拠点 OD ペアの脆弱度であり, W_{od}^u と W_{od}^s は平常時と災害 s 発生時の拠点 OD ペアの期待所要時間であり, C_{od}^u と C_{od}^s は平時と災害 s 発生時の拠点 OD ペアと修復しない経路集合である.

拠点 od 間の分散パラメータ θ_{od} は以下に定義する.

$$\theta_{od} = \frac{\pi}{\sqrt{6\lambda t_{od,sp}}} \quad (4)$$

ここで, λ は補助パラメータ, $t_{od,sp}$ は拠点 OD ペアの最短経路である. なお, 本検討では拠点間の経路探索はダイクストラ法を用いて行い, 主に最短経路と次善経路の2経路を経路集合とする.

(3) 最適整備順序の定式化

前節の防災機能評価手法を1つの指標として, 早期に対象ネットワークを強靱化するための整備順序の最適化を考える.

まず, 整備対象はハザードリンク $l_h \in \mathcal{L}_h (\mathcal{L}_h \in \mathcal{L})$ と新規整備対象リンク $l_n \in \mathcal{L}_n$ を合わせた整備対象リンク

$l_m \in \mathcal{L}_m = (\mathcal{L}_h + \mathcal{L}_n)$ とする. 選択の際には, その中のハザードリンク l_h を選択・整備した場合, リンク l_h 内の交通施設 \mathcal{F}_l の耐災害化整備が完了したとして l_h のハザードを解除する. 一方, 新規整備リンクを選択・整備した場合, リンク l_n の接続・耐災害整備が完了したとして新しくリンク集合 \mathcal{L} に追加する.

リンク整備に要する期間は, どのリンクも同じで1と仮定し, 整備対象リンクの本数分整備期間を設ける. そのため, 整備対象リンク本数が仮に $q(\mathcal{L}_h + \mathcal{L}_n)$ 本だった場合には総整備期間 r は $r = q(\mathcal{L}_h + \mathcal{L}_n)$ となる. しかし, 実際は整備費用に上限が定められているため $q(\mathcal{L}_h + \mathcal{L}_n)$ 本を一度に整備するのは不可能であると考えられる. そのため, 年間予算として C (億円)を定め, その上限に達するまで整備リンクを選択することにした. また, 各リンクの整備にかかる費用は, 全てリンク長 l_{length} に依存するとしてリンク長 l_{length} の値をそのまま整備費用 l_{cost} に充てた. ここで, 各リンク整備期間時刻のステップを t_{step} とし, 予算上限に達した時のステップを T とした場合, 一連の整備は $0 < t_{step} \leq T (T < r)$ の期間に行われることになる. この時の T は, リンクの方法によって変動し, 整備効果との兼ね合いであらゆるパターンが考えられる. また, 選択行動の終了条件には, ①費用が予算を上回った場合, ②強靱性が上限に達してこれ以上の改善が見込めない時, ③全ての整備対象リンクを整備し終わった場合の3状態を用意した. これらをまとめて図-1に示す.

防災機能評価は, 拠点間の経路単位で考えるため, 複数選択して初めて効果が発現する場合が存在する. 代替するリンクがそれよりも前のステップで整備されていた場合などに同じリンクでもどのステップで選択されるかによって強靱性の増加量が変動することが推測される.

このような特性の中で, 一般的には予算上限の範疇での総合的な強靱性の増加量の最大化は, どの順序で行っても最終的に到達するネットワーク状態が同じならばリン

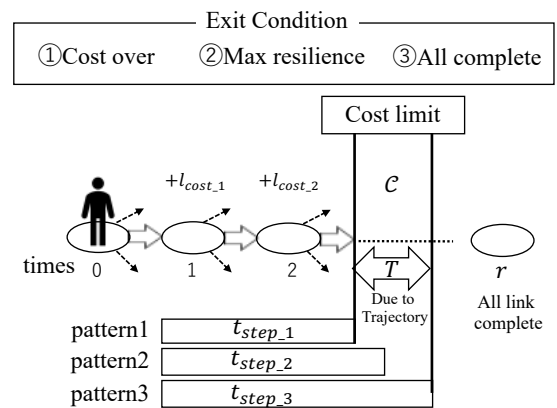


図-1 整備順序選択の構造

ク整備の順序によらないので、予算上限に達した時点 T での選択リンクの組み合わせ問題になってしまう。整備順序に優劣をつけるためには単なる整備効果の増加量によらない便益を考える必要がある。

そこで、災害の発生時期が不定期であることを考慮すれば、整備の効果発現を早期に出せることが重要であり、便益が高いと考えた。そのため、基本的には強靭度増加が早い段階で起こる選択順序パターンに高い便益を当てる方法を検討する。初めに、式(1)の脆弱度を1から減じて逆の意味を強靭度に変換した指標を考える。

$$K_{od} = 1 - Z_{od} \quad (5)$$

さらに、ネットワーク全体の強靭化を目指すため対象ネットワーク内の全各拠点ペアの合計をとり、合計強靭度 K_{tstep} と以下のように設定する。

$$K_{tstep} = \sum_{od \in \mathcal{H}_p} K_{od} \quad (6)$$

ここで、 \mathcal{H}_p :全拠点ペア集合(\mathcal{H} から設定) である。総和をとる理由は、各拠点間で使用する経路のリンクが被っている時、そのリンクを直した時の強靭化効果は複数の拠点ペアにまたがって影響を及ぼすため、合算することでネットワーク全体の性能評価指標とした。

次に、早期効果を考慮した順序評価として、式(7)に示す各ステップの強靭度を全整備対象リンクのステップ数分累積させた、強靭度推移グラフの面積を与える。早期に効果が出るものはそれ以降その効果が維持されるとして、それ以上の強靭度が蓄積されていくため遅くに効果が出るものより大きな評価が得られるはずである。また、選び方によっては数個のステップ数の違いで強靭度が逆転するパターンが十分起こり可能性があり、発現の早さと効果の大きさを天秤にかけのが難しいが、その場合も面積による評価ならば上記同様の理由から逆転する場合も十分に考慮した評価が可能である。

$$\max \begin{cases} \left(\sum_{tstep \in T} (K_{tstep}) l_{m,tstep-1} \right) + K_T(r - T) \quad \dots (a) \\ \left(\sum_{tstep \in T'} (K_{tstep}) l_{m,tstep-1} \right) + K_T(r - T') \quad \dots (b) \\ \sum_{tstep \in T=r} (K_{tstep}) l_{m,tstep-1} \quad \dots (c) \end{cases} \quad (7)$$

$$s. t.) \quad \sum_{tstep \in T} l_{m,tstep-1} = 1$$

$$\sum_{tstep \in T} l_{m,tstep-1}^{cost} \leq C$$

ここで、(a) は①費用が予算を上回った場合、(b)は②強靭性が上限に達してこれ以上の改善が見込めない時、(c) は③全ての整備対象リンクを整備し終わった場合最大の

強靭度に達した場合にそれぞれ該当する。なお、 T' は上限の強靭度に達した時のステップ数、 $l_{m,tstep-1}$ は時刻 $tstep - 1$ における任意のリンク l_m の整備有無、 $l_{m,tstep-1}^{cost}$ は整備費用である。

4. 深層強化学習を用いた提案モデルの解法

(1) 深層強化学習の概要

強化学習とは、行動主体が仮想的な環境(Environment)と相互作用を通して「一連の価値を最大化する行動」を学習する枠組みである。その相互関係は「行動主体(Agent)」、「状態(State)」、「行動(Action)」、「報酬(Reward)」の4つの要素からなる。マルコフ決定過程はこの相互依存を連鎖的に表すもので、 (S_t, A_t, P_t, R_t) により定義される。具体的には、 $s_t \in S_t$ を状態、 $a_t \in A_t$ を行動と呼び、 $P_t \rightarrow [0,1]$ は状態 s_t において行動 a_t をとった際に、次の状態 s_{t+1} に遷移する確率である。また $r_t = R(s_t, a_t) \in R_t$ は状態 s_t にて行動 a_t をとった際の報酬と定義され行動の評価に用いられる。行動主体は、この繰り返しの行動の中で将来に渡って得られる報酬の和が大きくなる様に行動する。また、この時状態 s_t において行動 a_t をとる確率を方策と称し、 $\pi(A_t|S_t)$ と表現する。 A_t は、時刻 t で取れる行動の全集合を表し、 S_t は時刻 t の状態の全集合を表す。なお、本稿では状態遷移確率 P_t は常に1と決定的にしており、整備の失敗や延期等は認めないこととした。

(2) 提案モデルの解法

前章で述べた整備順序最適化問題に強化学習を適用する際、提案モデルに応じた環境設計を行う必要がある。行動主体は、交通ネットワークの各時間ステップでの整備状況をもとに最適なリンクを選択していく「政府エージェント」とし、状態は、整備対象リンク \mathcal{L}_m の整備の有無を1(整備済み)または0(未整備)を用いて表したベクトルで定義する。行動は整備対象リンク \mathcal{L}_m のリンク番号を並べたベクトルとした。報酬は強化学習において式(7)をそのまま利用すると発散する挙動を示したため、過度に値が大きくなり、早期に高い効果を出すほど効果が高くなる様、式(16)を用いた。

$$r_{tstep} = \frac{K_{tstep} - K_0}{tstep} \quad (16)$$

$$\max \sum_{tstep \in T} (r_{tstep}) l_{m,tstep-1} \quad (17)$$

$$s. t.) \quad \sum_{tstep \in T} l_{m,tstep-1} = 1$$

$$\sum_{tstep \in T} l_{m,tstep-1}^{cost} \leq C$$

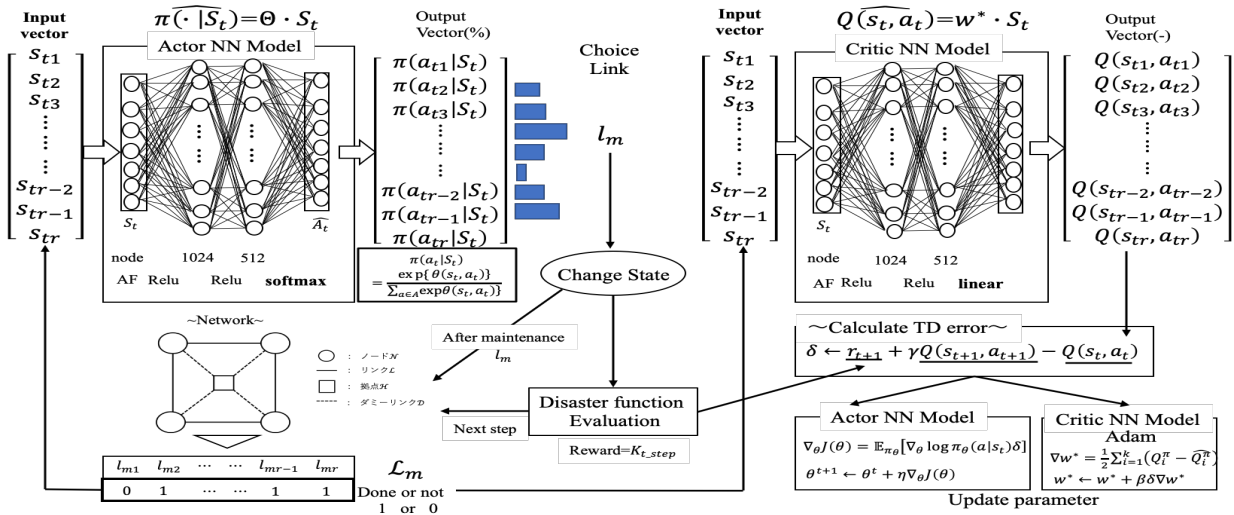


図-2 Actor-Critic アルゴリズムを用いた整備リンク最適整備順序決定モデルのフロー図

ここで、 K_0 は初期状態の合計強靱度である。各ステップに初期状態からの強靱度増加量を平均化し、整備効率の点数とみなして与える意味を持つ。本問題では、この式(16)の1エピソードの総和の最大化(式(17))を目指す。これにより多くのリンクを直し、早期に効果の高いものが出るほど大きな便益が見込める。

関数近似には、状態数の大きさから行動価値関数や方策関数の関数近似にはニューラルネットワーク(以下 NN)を利用した深層学習構造のモデルを作成した。本稿では Q-learning と Actor-Critic の二つのアルゴリズムを使用したがその違いは方策を確率関数にしているか否かのみで価値関数モデルの構造は同一にしたため、Actor-Critic モデルのみ全体の流れを図-2 に図示する。なお、図中の AF は活性化関数を示す。

5. テストネットワークへの適用と検証

(1) 小規模仮想ネットワークにおける検証

作成したモデルのパフォーマンスを測るため、全探索が可能な図-3の小規模仮想ネットワークを用いる。整備対象リンク数は、全探索可能な範囲内で最大の本数に近い6本と7本の場合を考えた。図-4 および表-1 に各学習結果および精度を示す。両手法とも学習回数に合わせて整備効果の収益が増加しており、TD 誤差がゼロに漸近する挙動を示している。精度に注目しても、両手法ともほぼ 100%の最適解を出していることが確認できた。これについては、6本の場合の720パターンに対して、2000回も学習を行えば最適解に到達するはずだが、7本の場合の5040パターンに対して、その半分以下の試行回数で最適解に到達したことから提案モデルの有効性が示された。また、計算時間に関しては、試行回数にも依存するため6、7本の場合は全探索が最速であるが、8本以上の

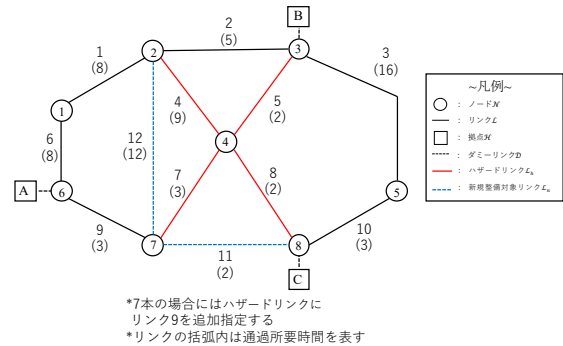


図-3 精度確認用小ネットワーク

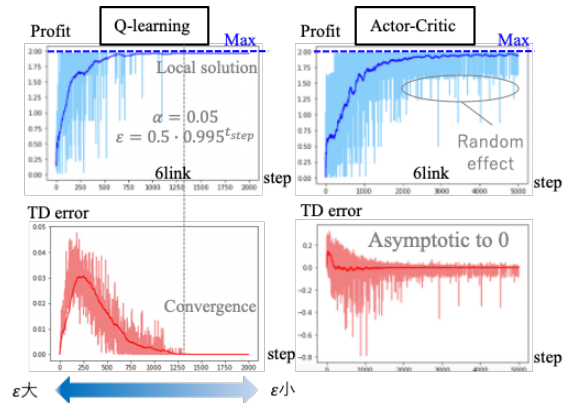


図-4 Q-learning と Actor-Critic の学習プロセス

表-1 モデル精度の結果

整備対象リンクが6本の場合			
	全探索	Q-learning	Actor-Critic
試行回数	720	2000	5000
計算時間(min)	5.21	36.99	38.54
合計報酬	1.98	1.96	1.98
決定順序	11→5→8→4→7	11→5→8→4→12	11→8→5→4→7
収束の有無	-	有	無
解の精度	100%	98.90%	100%
整備対象リンクが7本の場合			
	全探索	Q-learning	Actor-Critic
試行回数	5040	2000	5000
計算時間(min)	35.86	55.2	58.48
合計報酬	1.32	1.32	1.32
決定順序	5→8→9→11→7	5→8→9→11→7	5→8→9→11→7
収束の有無	-	有	無
解の精度	100%	100%	100%

場合では提案モデルの方が早いことが考えられる。

収束性については、Q-learning はどちらも学習回数約 1250 回で TD 誤差が 0 になり選択も固定されたことから収束が確認されている。一方で Actor-Critic の方は厳密な収束性の確認は不可能であるが、行動価値関数はランダム選択によるブレがあるものの強靱度の 0.1~3.3%の誤差まで TD 誤差が減少していることから、厳密な収束ではないが収束に近づいていると判断した。

以上より小規模な問題による検証において、Actor-Critic および Q-learning の精度は、概ね 100%を達成できるとが確認できた。しかしながら、探索の深さが十分ではないこともあり、探索初期から最適解に近い解をランダム選択で出しているため、規模を上げた場合の最適解が出せるのか疑問の余地が残る。

(2) Sioux Fallsネットワークにおける検証

全探索不可能な規模のネットワークを用いてモデルの有効性を検討する。対象ネットワークは米の Sioux Fall を用いる(図-5)。整備対象リンク $l_m \in \mathcal{L}_m$ は 49 本に設定し、予算は全リンクを整備する場合の 25%にした。これにより無駄なリンク整備を避けて重要なリンクを効率よく選択できるか挙動を検証する。適用した結果は図-6 の上図の様になる。Actor-Critic に関しては、学習回数の増加に伴って収益の増加が見られ、TD 誤差も継続的に減少して最終的には価値関数 NN は 0.3%~3%の幅に収まっている。一方で Q-learning は前節で発揮した効果は十分出でおらず、収束はしているものの局所的な解に陥っていることが確認できた。この理由は規模を大きくしたことによって選択順のパターン数が爆発的に増えたこと

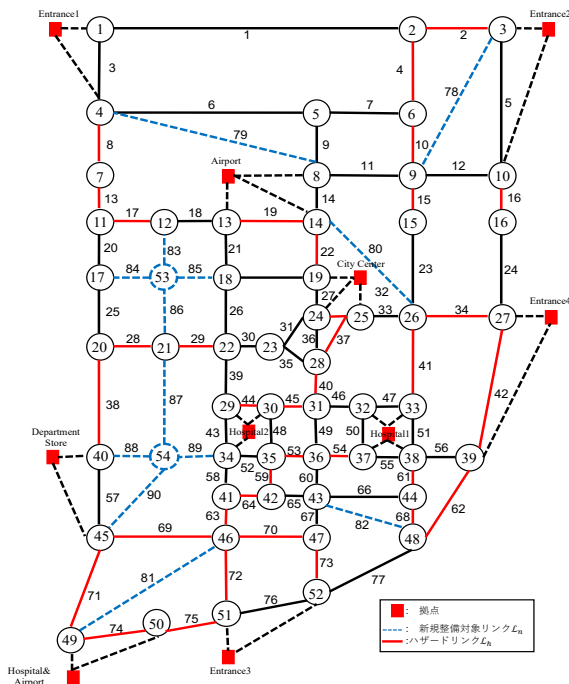


図-5 Sioux Falls ネットワーク

で探索が進みにくくなってしまったこととパラメータ更新幅が小さかったことが考えられる。

全拠点間合計強靱度の変化は図 10 の中図に示している。ここで AC は Actor-Critic で Q は Q-learning を示す。この推移を見て見ると、Q-learning モデルより Actor-Critic の方が最大で 39.9%増しで良いパフォーマンスを出していることがわかる。

以上の結果より、今回の問題においては Actor-Critic モデルの方が Q-learning よりも規模の大きい問題に適していると言える。ただし、その精度は全探索が行えないため確認が不可能である。さらに、強化学習は試行錯誤を通して経験を学習するため毎回学習挙動が異なり、再現性が乏しいことが指摘されている⁹⁾。

そこで今回は最後に計 5 回同じ Actor-Critic モデルで計算を行い、そのステップごとの強靱度のばらつきを確認した。結果は図-6 の下図の様になった。早期のステップの分散は明らかに小さく、7 ステップ目から広がっている様子が見て取れる。これは初期の方の探索パターンが小さく確実に良い選択を探索できるからと考えられる。この意味で Actor-Critic モデルは初期の選択は比較的效果が出やすいものを選べる可能性が高いと推察される。また、早期整備効果は式(7)より算定し、選択順序とともに

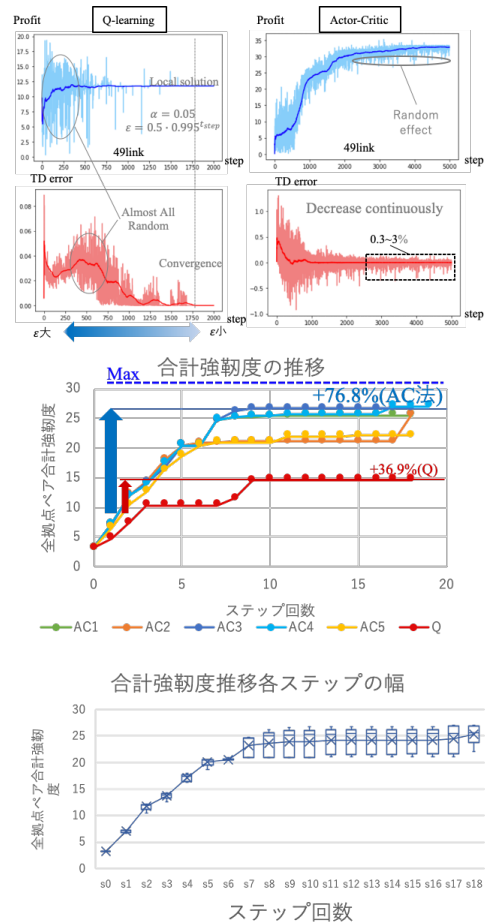


図-6 学習の様子(上), 合計強靱度推移(中), Actor-Critic モデルの再現性(下)

表-2 に示す。これを見ると初期の整備順序が完全に一致していることはないものの、選択リンクには、特に初頭のステップにて共通のものが散見される。これは早期に効果が発現できる同型のリンク順を決定できているからだと考えられる。また、全パターン of 予算超時の 19 ステップまでの整備早期効果を式(7)より算出した結果、最終強靱度と早期の効果発現の大きさが一致しないケースも確認された。このため、最終的な強靱度と予算内での早期の効果発現のどちらを優先すべきか更なる議論を行う必要がある。この点に関しては今後の課題としていきたい。

以上の結果より「全費用の約 25%の費用で、近似解として最大値の 61.1%~76.8%の増加が見込め、ステップ初期から効果の発現が期待できる」ことは、効率的な整備方針を立てる上でこのモデルが有用であると考えられる。

6. 本研究の成果と課題

本研究では、大規模災害による道路ネットワーク破壊の被害を最小限に抑えるため、限られた年間防災予算内で早期に効果が発現できるリンク整備順序の決定手法を提案した。Actor-Critic 法と Q-learning という強化学習手法を用いて、小規模なネットワークにてその精度の妥当性を示した後、Sioux Falls ネットワークに適用することでその有効性に言及した。

しかしながら、本検討にはいくつかの課題点が存在する。選択枝数の増加に伴い計算時間が爆発的に増加すること、より大規模なネットワークへの適用の検討ができていないことに関しては、計算時間の効率化を念頭に置くべきである。選択枝数によって指数関数的に探索領域が増加するため、計算機の高精度化や強化学習の並列処理等が必要になる可能性がある。大局的最適解ではなく、精度も大規模問題では必ずしも保証されないことについては、大規模な問題にそのまま適用した場合、完全な最適解を得ることは難しい。そのため「AlphaGo Zero」の様に探索効率を上げる工夫が必要となるだろう。局所解

表-2 Actor-Critic の結果 (5 回分)

	選択順序	最終合計強靱度	早期効果 (19ステップ まで)
AC_1	45,22,73,75,40,88,89,54,53,68,61,13,64,59,4,44,71	25.53	424.46
AC_2	22,40,73,75,45,54,61,68,44,4,59,89,63,13,8,53,64,90	25.61	378.73
AC_3	45,22,40,73,75,69,72,63,54,64,4,68,89,59,53	26.71	438.38
AC_4	45,22,73,75,40,90,89,54,53,61,68,64,13,4,63,59,71	27	429.98
AC_5	40,19,73,75,45,22,54,72,59,13,63,53,4,68,61,64,82	22.12	370.5

を対策する方法として、教師データを学ばせる逆強化学習が存在するがそもそも正解のない問題であるため、他手法との比較によって優劣をつけることしかできないため適用は難しいだろう。収束性が確認できる手法ではないことについては、災害の発生は確率的であることに対して今回は決定的に決めてしまっているため、災害発生パターンを複数考える必要がある。以上の改善点を示して今後のモデルの改良と実務への応用を期待する。

参考文献

- 1) 国土交通省道路局都市局：道路ネットワークの防災機能の向上効果計測マニュアル（案）, 2016.
- 2) Nagae, T., Fujihara, T., Asakura, Y.: Anti-Seismic Reinforcement Strategy for Urban Road Network, Transportation Research Part A, Vol. 45, No. 5, pp. 813–827, 2012.
- 3) 澁谷長史, 牧野貴樹：これからの強化学習, 森北出版, 2016.
- 4) Francisco S Melo, Sean P Meyn, M Isabel Ribeiro : An Analysis of Reinforcement Learning with Function Approximation, In Proceedings of the 25th international conference on Machine learning, pp.664-671.ACM, 2008
- 5) 石井信, 佐藤雅昭：統計的手法にもとづく強化学習と制御ルールの獲得, 計測と制御, Vol.39, No.12, pp.763-768, 2000.
- 6) 久保隆宏: Python で学ぶ強化学習 入門から実践まで[改定第2版], 講談社, 2019

(2020. 10. 2 受付)

OPTIMAL ROAD IMPROVEMENT PROCESS UNDER THE NATURAL DISASTER USING THE DEEP REINFORCEMENT LEARNING

Hideki YAGINUMA, Yu JINNO, Shintaro TERABE, and Kosuke TANAKA