

強化学習を用いた運搬ロボットの通信品質保証に関する 一検討

新宮 裕章¹ 藤橋 卓也¹ 工藤 理一² 高橋 馨子² 村上 友規³ 渡辺 尚¹ 猿渡 俊介¹

概要：携帯電話網や無線 LAN の進化によっていつでもどこでも無線ネットワークに接続できる環境が整備されたことで、Automatic Guided Vehicle (AGV) による部品や荷物の自動搬送、無人走行車やドローンの遠隔操作などネットワークを介してロボットを制御する応用が広がっている。ネットワークを介してロボットを制御する場合、ネットワーク資源を共有している多数のロボットそれぞれに対して安定的な制御を確立するために必要な通信品質を維持することが求められる。本稿では、ロボットのタスクを考慮してネットワークの制御の最適化を実現する強化学習のフレームワーク「CoRein」を提案する。CoRein は、ロボットの操作とネットワークの設定を行動、ロボットとネットワークで取得できる情報を状態、ロボットに与えられたタスクの達成状況とネットワーク性能を加味した報酬関数を設定して、行動価値関数を Deep Neural Network (DNN) で強化学習する。計算機シミュレーションを用いて複数台のロボットが定められた経路を移動して荷物を運搬する環境で評価した結果、CoRein を用いることで各ロボットに対して十分な通信速度を保証するネットワーク設定が可能となり、各ロボットが与えられたタスクを効率よく達成できることを確認できた。

1. はじめに

5G をはじめとした低遅延下を目指したワイヤレスネットワークの登場にあいまって、ネットワークを介してロボットを制御する需要が高まっている。例えば倉庫や工場では、無線ネットワークを介して制御された無人搬送車による貨物輸送の実証実験を通して業務の高度化・効率化が期待されている。他にも自動運転車や建設機械の遠隔操作、自律型作業ロボット、ドローン運送など、さまざまな施設や企業で無線ネットワークを介した遠隔制御を利用した実証実験が進められている。

無線ネットワークを介してロボットを制御する場合、周辺環境や他のロボットによる通信などの要因によってネットワーク性能が低下してロボットの動作に影響が生じることが予想される。より具体的には、ロボットと基地局との間の通信距離、マルチパスによるフェージング、ロボットの移動によるシャドウイング、他のロボットの通信との通信衝突などが発生する。最悪の場合、ロボットは無線ネットワークから切断される。無線ネットワークから切断されたロボットは制御を失って、ロボットを通して提供するサービスを継続できなくなる可能性がある。中央労働災害

防止協会によるメーカー向けの実地調査 [1] では、無線通信の断絶によって制御を失った無人搬送車と人との接触事故が報告されている。生産性の向上や安全性の担保のためには、ロボットとネットワークは常時接続されてサービスに必要な通信品質を満たすことが求められている。

サービスに必要な通信品質を担保するために、Received Signal Strength Indicator (RSSI) に基づくネットワークの最適化手法 [2-4] が検討されている。一方で、RSSI による最適化手法は他のロボットに起因するネットワーク性能の低下を把握できないため、各ロボットが十分なネットワーク性能を得られない可能性がある。十分なネットワーク性能が得られなかったロボットは安全のため速度を落とすなどして、与えられたタスクを十分に遂行できないことにもつながる。本稿ではネットワーク性能の最低品質を保証した手法を実現することで各ロボットにおけるタスクの遂行効率を向上することを目指す。

本稿では、深層強化学習を用いたネットワーク最適化手法「CoRein」を提案する。CoRein では、ネットワーク性能やロボットのタスク実行性能に応じてネットワークの設定やロボットの挙動を強化学習によって最適化する仕組みを提供する。具体的には、ロボットの操作とネットワークの設定を行動、ロボットで取得したセンサ情報やネットワークの設定情報を状態、ロボットに与えられたタスクの

¹ 大阪大学大学院情報科学研究科

² 日本電信電話株式会社 NTT 未来ねっと研究所

³ 日本電信電話株式会社 NTT アクセスサービスシステム研究所

達成状況やネットワーク性能を報酬関数として、行動価値関数を ϵ -greedy 法を用いて DNN で近似しながら最適化する強化学習を実現した。計算機シミュレーションによって提案手法「CoRein」を評価した結果、RSSIに基づいてアクセスポイントを選択する naive な手法に比べて、最低通信品質を満たした割合を最大で 67 ポイント向上させつつタスク達成率を 60 ポイント向上できることが確認できた。

本稿の構成は以下の通りである。2 節では、ロボットとネットワーク制御の関連研究について述べる。3 節では、本稿で想定する貨物運搬モデルについて述べる。4 節では、提案手法である CoRein について述べる。5 節では、評価結果について述べる。最後に 6 節では、結論を述べる。

2. 関連研究

2.1 ネットワーク最適化

本研究は、ネットワーク最適化に関連する。ネットワーク最適化の研究としては、主に基地局配置設計手法、基地局選択手法、リソース制御手法が挙げられる。基地局選択手法としては、RSSI に基づいて基地局を選択するなどの端末側が主導となって基地局を選択する場合もあるが、本研究では、端末の基地局の選択自体もネットワークの最適化であると位置づける点に注意されたい。

無線基地局の配置を設計する研究としては、平均スループットを向上させるために基地局を配置する研究 [5–8]、平均アクセス遅延を軽減するために基地局を配置する研究 [9]、カバレッジを増加させるために基地局を配置する研究 [10–12] などが挙げられる。文献 [5] では、深層強化学習を用いてデータ収集などのサービスを地上に提供する Unmanned Aerial Vehicle (UAV) を都市環境において最適に配置する手法を提案している。文献 [6] では、ユーザの位置が時間的に変化することを考慮して、協調的確率的近似に基づいて Access Point (AP) の位置とカバレッジ半径を共同で最適化する手法を提案している。文献 [9] では、モバイル IoT デバイスとそのトラフィック負荷に動的に適應できるランキングベースの近距離最適配置アルゴリズムを用いた RNOPA を提案している。文献 [11] では、屋内測位精度とサービス範囲の向上を目的とした、多目的進化計算 P-Estra の有効性が検討されている。

基地局の選択は、携帯電話網や無線 LAN においてセルを移動するときのハンドオーバーの際に行われる [13]。ハンドオーバーでは、進行中のセッションの中断を最小限に抑えて通信を継続することが第一の目標となる。ハンドオーバー中の最大の遅延は、切り替え先の無線基地局を見つけるために必要なスキャン段階である [14]。スキャンを容易にするための技術として、様々なデータリンク層情報を上位相に提供する IEEE 802.11k [15]、Fast basic service set Transition (FT) によって無線基地局切り替え時に高

速で認証する仕組みを提供する 802.11r [16] などの規格が提案されている [17]。現状ではほとんどのハンドオーバーは RSSI に依存している [18]。RSSI だけでなく、消費電力や遅延などの要因に基づいて基地局選択をする仕組みも検討されている。例えば、ゲーム理論に基づいたネットワーク接続に関する研究として、ゲーム内のプレイヤーがユーザ対ユーザのもの [19] と、ユーザ対ネットワークのもの [20]、ネットワーク対ネットワークのもの [21] などが検討されている。また、文献 [22] では、動的マルチチャネルアクセス問題における未知のダイナミクスと法外な計算量という課題を克服するために、強化学習を利用したアプローチが提案されている。

接続端末に対する適切なリソース割当てに関する研究としては、適応的に周波数を割り当てる研究 [23]、適応的に電力を割り当てる研究 [24, 25]、アンテナを割り当てる研究 [26–28] などが挙げられる。例えば、文献 [23] では、十分に利用されていない周波数帯に動的にアクセスすることを可能にすることを目的として、統計的チャネル割り当てを用いた SCA-MAC プロトコルを提案している。文献 [29] では、セルラーネットワークにおけるエネルギー効率の向上を目的として、新しいマルチホップコグニティブセルラーネットワーク (MC2N) アーキテクチャを提案している。文献 [26] では、電力配分の設計とアップリンクビームフォーミング設計を分離することで、複雑な演算を必要としない深層学習に基づく高速ビームフォーミング設計手法を提案している。文献 [27] では、無線基地局によるビームフォーミングとインテリジェントサーフェスによる反射ビームフォーミングを共同で最適化することで、AP の送信電力を最小化する問題を定式化している。文献 [30] では、あるホストから別のホストまでのネットワーク経路をリスト表示する traceroute 機構と標準的なスイッチ機能を用いることで、通信経路の輻輳状態を学習するトラフィック負荷分散のフレームワークを検討している。

2.2 移動ロボットの経路計画

移動ロボットの経路計画については、これまでもいくつかの方法論 [31–33] が提案・報告されている。A*アルゴリズム [31] はヒューリスティックな経路探索アルゴリズムの中で最も使用されているアルゴリズムである。A*アルゴリズムは、探索の方針を決定するヒューリスティック関数 $h(n)$ から最適経路を選択するアルゴリズムである。ヒューリスティック関数 $h(n)$ は、スタートノードから n ノードまでの最小コストの推定値と、 n ノードからゴールノードまでの最小コストの推定値を足し合わせた関数である。ゴールまでの推定コストを用いることで最短経路の探索が常に目標地点の方向に進むことが保証されるため、効率的に経路探索を行うことができる。ヒューリスティックな経路探

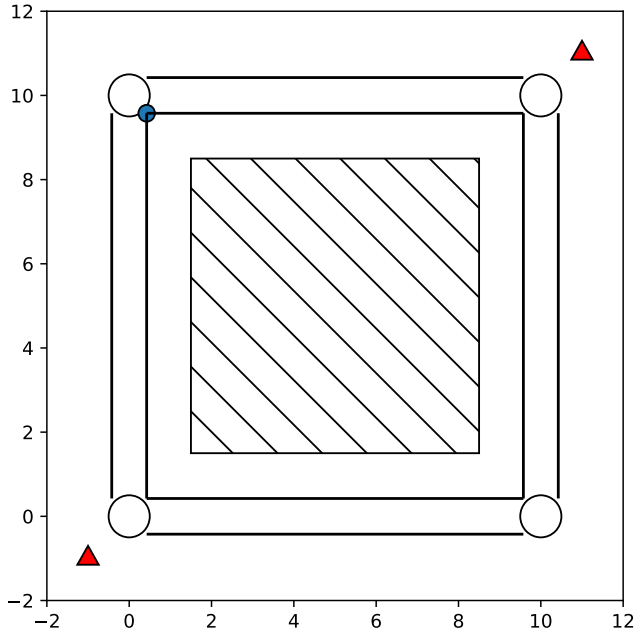


図 1: 貨物運搬モデルのマップ

探索アルゴリズム以外に、機械学習による学習ベースの経路探索アルゴリズム [32, 33] も検討されている。文献 [33] では、強化学習ベースの探索アルゴリズムに A* アルゴリズムを組み込むことで経路探索モデルの学習の収束を早めることに成功している。文献 [32] では、構造の類似した多くのロボットで学習した複数の経路決定モデルを異なる環境や利用形態にまたがって汎化する General Navigation Model (GNM) を提案している。

3. 貨物運搬モデル

本稿では、 $14 \times 14 \text{ m}^2$ の正方形の 2 次元平面環境を任意の台数のロボットが移動して貨物運搬を行う計算機シミュレータを実装した。図 1 に想定する貨物運搬モデルを示す。深層強化学習のフレームワークを適用するために、ロボットの動作は簡略化したモデルであることに注意されたい。以降、横軸方向 X 地点、縦軸方向 Y 地点の位置座標を (X, Y) として示す。

モデルは、4 つの中継地点、8 本の経路、1 つの障害物、任意の台数のロボット、2 台の無線 AP で構成される。中継地点は、荷物の積み込み地点と積み下ろし地点として指定される地点である。中継地点は $(0, 0)$, $(10, 0)$, $(0, 10)$, $(10, 10)$ の 4 つの地点に設定した。中継地点は半径 50 cm の円形とした。経路は中継地点を始点と終点にもつ線分である。縦軸方向と横軸方向に隣接する中継地点どうしは 2 本の経路で結ばれている。マップ中央から見て外側の経路は時計回り方向にロボットが進行する経路、内側の経路は反時計回りにロボットが進行する経路とした。2 次元平面上の中央には 1 辺 7 m の正方形の障害物が設置され

ている。障害物は電波を遮蔽する役割を持つ。無線 AP は $(-1, -1)$ 地点と $(11, 11)$ 地点の 2 か所に設置した。ロボットは、無人搬送車のような移動する機体を意味している。ロボットは積み込み地点で荷物を受け取り、積み下ろし地点として指定された中継地点まで経路上を移動する。ただし、同じ経路上のロボット同士は前後関係を入れ替えることができないものとした。積み下ろし地点まで荷物を運搬したロボットはその場で次の荷物を受け取って、別の中継地点に積み下ろし地点を指定されるものとした。各ロボットは 1 秒ごとに接続する無線 AP を選択するものとした。各ロボットは利用可能な 2 つの無線 AP のうち、いずれかの無線 AP を選択して接続し、自身が保持するデータを無線チャンネルを介してアップロードするものとした。このとき、各ロボットから接続先無線 AP に対する上り方向の通信スループットは同一の無線 AP に接続するロボット数にしたがって定まるものとした。同一の無線 AP に接続するロボット数が複数の場合、通信に利用できる帯域は等分配されると仮定した。他のロボットの通信も考慮した、ロボット i から接続先無線 AP m に対する上り方向の通信スループット $C_{i,m}$ [bps] を次の通り定めた。

$$C_{i,m} = \frac{B \log_2 \left(1 + \frac{S_{i,m}}{N} \right)}{L_m} \quad (1)$$

ここで、 B は帯域幅 [Hz]、 N はノイズフロア [W]、 L_m は無線 AP m に接続されるロボットの台数である。帯域幅 B を周波数 2.4GHz の Wi-Fi の帯域幅である 20MHz、ノイズフロア N を -90 dBm と定めた。送信電力 $S_{i,m}$ は以下の通り定めた。

$$S_{i,m} = P_i = \alpha^2 10^{\frac{\beta}{10}} g(d_{i,m}) P_t G_t G_r \quad (2)$$

P_i はロボット i の送信電力 [W]、 $d_{i,m}$ はロボット i と無線 AP m 間の通信距離 [m]、 G_t はロボットのアンテナ利得、 G_r は無線 AP のアンテナ利得である。ロボット i の送信電力 P_i を一般的な無線 LAN モジュールの送信電力にしたがって 1mW、ロボットのアンテナ利得 G_t を 1、無線 AP のアンテナ利得 G_r を 1 と定めた。レイリーフェージング係数 α^2 はレイリー分布に従うものとして以下の通り定めた。

$$\alpha^2 \sim x \exp \left(-\frac{x^2}{2} \right) \quad (3)$$

シャドウフェージング係数 $10^{\frac{\beta}{10}}$ は電力分布特性が対数正規分布 (Log-Normal) で近似できるため、 β は以下の通り定めた。

$$\beta \sim \mathcal{N}(0, 6) \quad (4)$$

$\mathcal{N}(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 にしたがう正規分布である。経路損失関数 $g(d_{i,m})$ は以下の通り定めた。

$$g(d_{i,m}) = \frac{\lambda^2}{16\pi^2 d_{i,m}^n} \quad (5)$$

λ は送信電波の波長 [m] である．送信電波の波長 λ を 2.4GHz の電磁波の波長 0.124m と定めた．経路損失指数 n_e は以下の通り定めた．

$$n_e = \begin{cases} 2 & (\text{if } e = \text{LOS}) \\ 4 & (\text{else if } e = \text{NLOS}) \end{cases} \quad (6)$$

e はロボットと無線 AP を結ぶ直線上に障害物が存在しない場合は LOS，障害物が存在する場合は NLOS をとるパラメータである．

十分なネットワーク性能が得られなかった場合，ロボットは安全のため強制的に速度を落とすように制御されるものとした．本稿で想定する貨物運搬モデルでは，ロボットがサービスを継続するための最低通信品質を 100 Mbps とした．上り方向のスループットが 100 Mbps 以下の場合，ロボットはいかなる指令を受けていた場合でも，移動速度を 0.1 m/s に設定されるものとした．

4. 提案手法：CoRein

本稿では，強化学習によってロボットのタスクを考慮してネットワークの制御の最適化を実現する「CoRein」を提案する．CoRein は，ロボット部，ネットワーク部，ロボット・ネットワーク制御部の 3 つの要素から構成される．ロボット部は，ロボット・ネットワーク制御部からコマンドを受け取って物理空間において任意のタスクを実行する．ネットワーク部は，ロボット部とロボット・ネットワーク制御部を接続する．ロボット・ネットワーク制御部は，各ロボット部とネットワーク部から情報を収集して，ロボット部に対して取るべき挙動や，ネットワーク部に対して接続する無線アクセスポイントなどの通信設定の指令を出す．各ロボット部からの情報は，ロボットの位置・向きなどの状態を意味する．ネットワーク部からの情報は，RSSI などのロボットに具備された無線 LAN インターフェースから取得できるネットワークの情報や，無線アクセスポイントに設定されているチャンネル，スループット，パケットロス率，遅延などの情報を意味する．

4.1 ロボット部

ロボット部は，制御対象となる AGV，ドローン，作業車両，自動運転車などのロボット自体を意味している．ロボットに搭載されているネットワークインターフェースの設定に関しては CoRein ではネットワーク部に属することに注意されたい．ロボット部は，入力としてロボットの操作，出力としてロボットの状態情報を持つ．ロボット i の操作は以下の式で表される．

$$B^i \triangleq \{b^{1,i}, b^{2,i}, \dots, b^{N_B,i}\} \quad (7)$$

N_B はロボットの操作の種類数である． $b^{n,i}$ はロボットで可能な操作と対応している．例えば，1 m/s で前進，停

止，右/左に 5 度回転などが相当する．ロボットの状態は以下の式で表される．

$$U^i \triangleq \{u^{1,i}, u^{2,i}, \dots, u^{N_U,i}\} \quad (8)$$

N_U はロボットの状態情報の種類の数である． $u^{n,i}$ に割り当てられた数字はロボットの各状態と対応している．より具体的には，ロボット自身が持つセンサで取得可能な情報を意味する．例えば，ロボットの位置情報，向き，バッテリー残量などが相当する．

4.2 ネットワーク部

ネットワーク部は，ロボット部とロボット・ネットワーク制御部を接続する機構である．無線ネットワークや有線ネットワークを想定している．繰り返しになるが，注意されたいのは，ロボットに具備された無線 LAN インターフェースはネットワーク部に属することである．ネットワーク部は，入力として通信設定，出力として通信情報を持つ．ネットワークの通信設定は以下の式で表される．

$$C \triangleq \{c^1, c^2, \dots, c^{N_C}\} \quad (9)$$

N_C はネットワークの通信設定の種類数である． c^n はネットワークに設定可能な項目と対応している．例えば，どのロボットがどの AP に接続するか，各 AP の周波数チャンネル，Software Defined Networks (SDN) を用いた場合には有線ネットワークのトポロジなどが相当する．ネットワークの状態は以下の式で表される．

$$V \triangleq \{v^1, v^2, \dots, v^{N_V}\} \quad (10)$$

N_V はネットワークの状態の種類数である． v^n はネットワークの各状態と対応している．より具体的にはネットワーク機器でモニタリングして取得可能な情報を意味する．例えば，ネットワーク全体の総スループット，パケットロス率，最低スループット，遅延などが相当する．

4.3 ロボット・ネットワーク制御部

ロボット・ネットワーク制御部は，各ロボット部に対して取るべき挙動や，ネットワーク部に対して接続する無線 AP などの通信設定の指令を出す機構である．ロボット・ネットワーク制御部は，入力として各ロボット部の状態情報 U^i と，ネットワーク部の通信情報 V ，出力として各ロボットへの操作 B^i と，ネットワーク部の通信設定 C を持つ．また，これらの値，ロボットに与えられたタスク，ネットワークの性能から以下の式で報酬が算出される．

$$R \triangleq \{r \in \mathbb{R} : r = g(s, a), s \in S, a \in \mathcal{A}\} \quad (11)$$

ここで行動 \mathcal{A} は，各ロボットのとるべき挙動 B^i ，ロボットの台数 M ，ネットワーク部の通信設定 C を用いて以下の式で表される．

$$\mathcal{A} \triangleq \left(\prod_{i=1}^M B^i \right) \times \mathcal{C} \quad (12)$$

また、状態 \mathcal{S} は、各ロボットの状態 U^i 、ロボットの台数 M 、ネットワーク部の状態 \mathcal{V} を用いて以下の式で表される。

$$\mathcal{S} \triangleq \left(\prod_{i=1}^M U^i \right) \times \mathcal{V} \quad (13)$$

また、 $g(s, a)$ はロボットに与えられたタスクやネットワークの状態によってアプリケーション毎に定義する報酬関数である。

ここである状態 s のときに行動 a を取ったときの報酬の推定値を返す行動価値関数 $Q(s, a)$ を考える。CoRein では、行動価値関数 $Q(s, a)$ を DNN で近似する。図 2 に CoRein で用いている DNN を示す。入力層はロボットとネットワークの状態情報 \mathcal{S} である。隠れ層は、各層は 128 ノードの全結合層 (Fully Connected (FC)), Rectified Linear Unit (ReLU) 活性化層の組を 3 回繰り返したものである。出力層は行動 \mathcal{A} を選択した場合の報酬の推定値である。

この DNN のパラメータを θ とする。CoRein では、各時刻 t において、以下の Huber 損失関数 $L(\theta_{t-1})$ を最小化するように θ の更新を行う。

$$L(\theta_{t-1}) = \begin{cases} \frac{1}{2} f(t-1)^2 & (\text{if } |f(t-1)| \leq 1) \\ |f(t-1)| - \frac{1}{2} & (\text{otherwise}) \end{cases} \quad (14)$$

ここで関数 $f(t)$ は以下の式で表される。

$$f(t) = r_t - Q(s_t, a_t | \theta_t) \quad (15)$$

式 (14) の最小化には Adam を用いることで以下の式で θ を更新する。

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) \nabla_{\theta_{t-1}} L(\theta_{t-1}) \quad (16)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) (\nabla_{\theta_{t-1}} L(\theta_{t-1}))^2 \quad (17)$$

$$\theta_t = \theta_{t-1} - \alpha \frac{v_t}{\sqrt{s_t + \gamma}} \quad (18)$$

α は学習率、 β_1 、 β_2 は平滑化定数、 γ はゼロ除算を防ぐための設定パラメータである。

また、行動選択の際には ϵ -greedy 法を用いることで行動選択にランダム性を持たせることで局所的な行動に陥らないようにしている。具体的には、ある時刻 t において、以下の式で行動 a を選択する。

$$a = \begin{cases} \arg \max_a Q(s_t, a | \theta_t) & (\text{with probability } 1 - \epsilon) \\ \text{randomly select } a \in \mathcal{A} & (\text{with probability } \epsilon) \end{cases}$$

この時、 ϵ は以下の式で算出する。

$$\epsilon = \epsilon_{end} + (\epsilon_{start} - \epsilon_{end}) \exp\left(-\frac{t}{T_{decay}}\right) \quad (19)$$

ϵ_{start} は ϵ の初期値、 ϵ_{end} は ϵ の収束値、 T_{decay} は ϵ が ϵ_{start} から ϵ_{end} まで変化する速さを表すパラメータである。

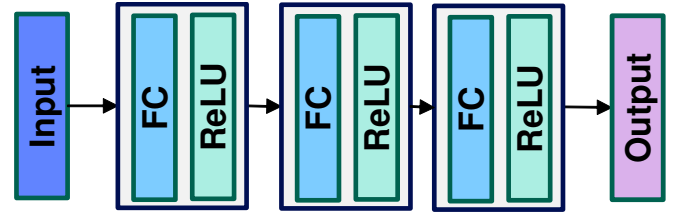


図 2: DNN の構造

表 1: パラメータの対応表

Hyperparam	CoRein
B^i	{1 m/s}
U^i	{X 座標, Y 座標, 速さ}
\mathcal{C}	{各ロボットの接続先 AP の指定}
\mathcal{V}	{各ロボットが接続している無線 AP}
M	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
$g(s, a)$	$\sum_{i=1}^M u^{3,i}$
α	0.01
β	0.99
γ	$1.0 \cdot 10^{-8}$
ϵ_{start}	1.0
ϵ_{end}	0.0
T_{decay}	100
P_t	0.001
λ	0.124
B	$20 \cdot 10^6$
N	$1.0 \cdot 10^{-12}$

5. 評価

5.1 評価環境

図 1 に本稿で想定する環境を再度示す。本評価では深層強化学習モデルの作成のために、ロボットが動作を開始してから 500 秒経過するまでを 1 エピソードとして 1000 エピソード分の訓練を行った。環境内のロボットの台数が学習モデルの性能に与える影響を評価するために、1 台から 10 台までのロボット台数で評価を行った。ロボットの初期位置は、(0, 10) の中継地点から (0, 0) の中継地点に向かう経路上に設定されるものとした。シミュレーション開始時、全ロボットの荷物の積み下ろし地点は (10, 0) の中継地点とした。その後、各ロボットの積み下ろし地点は、(0, 10) の中継地点と (10, 0) の中継地点に交互に設定されるものとした。ロボットが最短経路で荷物を積み下ろし地点まで運搬するために、経路する中継地点の決定には A* アルゴリズム [31] を採用した。ロボットの上り方向のスループットが 100 Mbps 以下の場合、ロボットの速さは 0.1 m/s に設定されるものとした。各ロボットのスループットが最低通信品質である 100 Mbps をどれだけ上回ったかを評価するために最低通信品質保障の観点で評価を行った。次に、結果的に全てのロボットによってどれだけ荷物が運搬されたかを評価するためにタスク達成割合の観点で評価を行った。

表 1 に提案手法「CoRein」の実装に使用したパラメータを示す．式 (7) で定義しているロボット i の操作 B^i を構成する要素 $b^{n,i}$ は, $b^{1,i}$ を 1.0 m/s で前進とした．式 (8) で定義しているロボット i の状態 U^i を構成する要素 $u^{n,i}$ は, $u^{1,i}$ をロボット i の X 座標, $u^{2,i}$ をロボット i の Y 座標, $u^{3,i}$ をロボット i の速さとした．式 (9) で定義しているネットワークの通信設定 C を構成する要素 c^n は, c^1 を各ロボットの接続先無線 AP の指定とした．式 (10) で定義しているネットワークの状態 \mathcal{V} を構成する要素 v^n は, v^1 を各ロボットが接続している無線 AP の情報とした．式 (11)(12)(13) で定義しているロボットの台数 M は 1 台から 10 台とした． $g(s, a)$ は全ロボットの移動の速さの合計として以下の通り定めた．

$$g(s, a) = \sum_{i=1}^M u^{3,i} \quad (20)$$

5.2 最低通信品質保証

図 3 にロボット台数に対する最低通信品質を満たした割合を示す．初期状態のロボットの位置関係が評価結果に与える影響を小さくするために, 評価する時間的な範囲は全てのロボットが最初の積み下ろし地点である (10, 0) の中継地点に到着した後からエピソード終了時までとした．横軸は, 環境内で動作するロボットの数を示す．縦軸は, 上り方向のスループットが閾値 0.1 Gbps を上回った割合を示す．各ロボットが無線 AP から得られた RSSI に基づいて接続先の無線 AP を選択する手法を Naive, 提案手法を CoRein とした．

図 3 の評価結果からは, 提案手法の「CoRein」は比較手法の「Naive」と比較して同等かそれ以上の性能を達成していることが分かった．具体的にそれぞれの最低通信品質を満たした割合は, ロボットの台数が 1 台の場合, Naive 手法が 100 % に対し CoRein は 98 %, ロボットの台数が 2 台の場合, Naive 手法が 100 % に対し CoRein は 98 %, ロボットの台数が 3 台の場合, Naive 手法が 100 % に対し CoRein は 98 %, ロボットの台数が 4 台の場合, Naive 手法が 76 % に対し CoRein は 98 %, ロボットの台数が 5 台の場合, Naive 手法が 25 % に対し CoRein は 92 %, ロボットの台数が 6 台の場合, Naive 手法が 5 % に対し CoRein は 70 %, ロボットの台数が 7 台の場合, Naive 手法が 4 % に対し CoRein は 41 %, ロボットの台数が 8 台の場合, Naive 手法が 4 % に対し CoRein は 36 %, ロボットの台数が 9 台の場合, Naive 手法が 4 % に対し CoRein は 30 %, ロボットの台数が 10 台の場合, Naive 手法が 4 % に対し CoRein は 25 % であった．

本稿で想定した環境ではロボットの数が 3 台以下であれば, 1 台の AP に全てのロボットが接続した場合でも各ロボットのスループットは最低通信品質である 100 Mbps

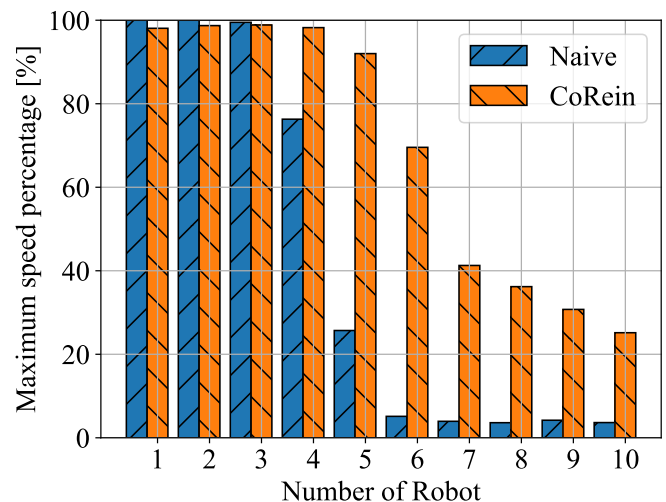


図 3: ロボット台数に対する最低通信品質を満たした割合

以下を下回らない．環境内のロボット台数が 4 台以上の場合, RSSI に基づいて接続先の無線 AP を選択する手法では, 1 台の無線 AP に 4 台以上のロボットが接続する状況が発生して各ロボットのスループットが最低通信品質である 100 Mbps を下回る割合が増加した．ロボットが 4 台以上の環境で Naive 手法の性能が低下しているのは, 単純な RSSI に基づいた接続手法では 1 台の AP に接続するロボットの台数を調整することができないために, スループットが急激に低下することが原因であると考えられる．一方で CoRein 手法は, 1 台の無線 AP に接続が集中することでスループットが低下することを学習して, 状況に応じて RSSI が低い方の無線 AP に接続する行動を選択するようになったため, 通信品質の低下を抑制できていると考えられる．

5.3 タスク達成割合

図 4 にロボットの台数に対するタスク達成割合を示す．横軸に環境内で動作するロボットの数, 縦軸にタスク達成割合を示す．タスク達成割合とは, 500 秒の運行時間で全てのロボットによって運搬された荷物数を, 全てのロボットが常に 1.0 m/s で移動した場合に運搬される荷物数で割った割合を意味する．ロボットが常に 1.0 m/s で動作し続けた場合, 1 つの荷物を運搬するために 20 秒経過するため, 500 秒でロボット 1 台あたり 25 回タスクを達成することができる．

図 4 の評価結果からは, 提案手法は RSSI に基づいて接続先の無線 AP を選択する手法と同等かそれ以上のタスク数を達成することが分かった．具体的にそれぞれのタスク達成割合は, ロボットの台数が 1 台から 3 台の場合, Naive 手法が 100 % に対し CoRein は 100 %, ロボットの台数が 4 台の場合, Naive 手法が 80 % に対し CoRein は 99 %, ロボットの台数が 5 台の場合, Naive 手法が 32 % に対し

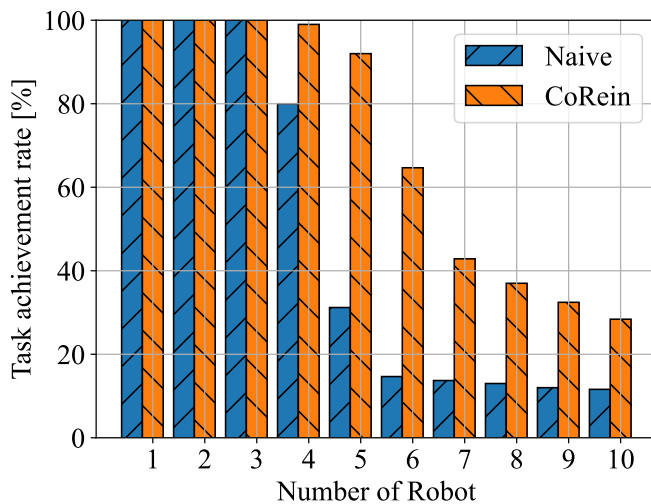


図 4: ロボット台数に対するタスク達成割合

CoRein は 92 % , ロボットの台数が 6 台の場合 , Naive 手法が 14 % に対し CoRein は 64 % , ロボットの台数が 7 台の場合 , Naive 手法が 13 % に対し CoRein は 42 % , ロボットの台数が 8 台の場合 , Naive 手法が 13 % に対し CoRein は 37 % , ロボットの台数が 9 台の場合 , Naive 手法が 12 % に対し CoRein は 37 % , ロボットの台数が 10 台の場合 , Naive 手法が 12 % に対し CoRein は 32 % であった .

図 3 の評価結果と同様に , Naive 手法のタスク達成割合が , ロボットが 4 台以上の環境で 80 % 以下に低下しているのは , 1 台の無線 AP に接続が集中することによる通信品質の低下に伴って , ロボットが 1.0 m/s の速さを保つことができないためであると考えられる . 一方で , CoRein 手法はロボットの移動速度をなるべく 1.0 m/s に保つために , 各ロボットのスループットが最低通信品質である 100 Mbps を下回らないような通信設定を学習したと考えられる . 具体的には , CoRein は 2 台の無線 AP にロボットからの接続を分散させるような通信制御を学習した結果 , Naive 手法よりもロボットの平均的な移動速度を向上させたことで優れたタスク遂行性能を達成できたと考えられる .

6. おわりに

本稿では , ネットワークを介したロボットの安定的な制御の実現にむけて , 深層強化学習を用いたネットワーク最適化手法「CoRein」を提案した . 具体的には , ロボットの操作とネットワークの設定を行動 , ロボットで取得したセンサ情報やネットワークの設定情報を状態として , ロボットに与えられたタスクやネットワーク性能を加味した報酬関数を設定することでネットワークの最適化制御に深層強化学習を適用した . 評価結果から , RSSI に基づいてアクセスポイントを選択しながら各ロボットが最短経路を選択する naïve な手法に比べて , 最低通信品質を満たした割合を最大で 67 ポイント向上させつつタスク達成率を 60 ポイ

ント向上させることが確認できた .

謝辞

本研究は JSPS 科研費 (JP19H01101 , JP22H03582) , NTT 研究所の支援の下で行った .

参考文献

- [1] 中央労働災害防止協会: 令和 2 年度厚生労働省委託技術革新に対応した機械設備の安全対策の推進事業 報告書 (2021).
- [2] Hernandez-Martinez, E., González-Sierra, J., Alvarez-Guzman, E., Fernandez-Anaya, G., Ferreira-Vazquez, E. D. and Flores-Godoy, J.-J.: Multi-robot formation based on RSSI power level and radiation pattern, *International Journal of Systems Science*, Vol. 53, No. 3, pp. 634–651 (2022).
- [3] Luthy, K. A., Grant, E. and Henderson, T. C.: Leveraging rssi for robotic repair of disconnected wireless sensor networks, *Proceedings IEEE International Conference on Robotics and Automation*, pp. 3659–3664 (2007).
- [4] Komatsu, T., Ohkubo, T., Kobayashi, K., Watanabe, K. and Kurihara, Y.: A study of rssi-based formation control algorithm for multiple mobile robots, *Proceedings of SICE Annual Conference*, IEEE, pp. 1127–1130 (2010).
- [5] Krijestorac, E., Hanna, S. and Cabric, D.: UAV Access Point Placement for Connectivity to a User with Unknown Location Using Deep RL, *IEEE Globecom Workshops*, pp. 1–6 (2019).
- [6] Zhang, Y. and Dai, L.: Joint optimization of placement and coverage of access points for IEEE 802.11 networks, *IEEE International Conference on Communications (ICC)*, pp. 1–7 (2020).
- [7] Dastgheib, M. A., Beyranvand, H. and Salehi, J. A.: Optimal placement of access points in cellular visible light communication networks: An adaptive gradient projection method, *IEEE Transactions on Wireless Communications*, Vol. 19, No. 10, pp. 6813–6825 (2020).
- [8] Zhou, B., Tu, W., Mai, K., Xue, W., Ma, W. and Li, Q.: A novel access point placement method for WiFi fingerprinting considering existing APs, *IEEE Wireless Communications Letters*, Vol. 9, No. 11, pp. 1799–1802 (2020).
- [9] Zhao, L., Sun, W., Shi, Y. and Liu, J.: Optimal placement of cloudlets for access delay minimization in SDN-based Internet of Things networks, *IEEE Internet of Things Journal*, Vol. 5, No. 2, pp. 1334–1344 (2018).
- [10] Qiu, S., Chu, X., Leung, Y.-W. and Ng, J. K. Y.: Joint Access Point Placement and Power-Channel-Resource-Unit Assignment for 802.11 ax-Based Dense WiFi with QoS Requirements, *IEEE Conference on Computer Communications*, pp. 2569–2578 (2020).
- [11] Korbel, P., Hausman, S. and Barba, P. D.: Application of Evolutionary Approach for Multi-objective Improvement of Indoor Access Point Placement, *2019 13th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–4 (2019).
- [12] Alsmady, A. and Awad, F.: Optimal Wi-Fi access point placement for RSSI-based indoor localization using genetic algorithm, *8th International Conference on Information and Communication Systems (ICICS)*, pp. 287–291 (2017).
- [13] Tekinay, S. and Jabbari, B.: Handover and Channel Assignment in Mobile Cellular Networks, *IEEE Communi-*

- cations Magazine*, Vol. 29, No. 11, pp. 42–46 (1991).
- [14] Mishra, A., Shin, M. and Arbaugh, W.: An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process, *SIGCOMM Computer Communication Review*, Vol. 33, pp. 93–102 (2003).
- [15] IEEE Std 802.11k-2008: IEEE Standard for Information Technology - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Fast Basic Service Set (BSS) Transition, *Amendment to IEEE Std 802.11-2007*, pp. 1–244 (2008).
- [16] 802.11r 2008, I. S.: IEEE Standard for Information Technology - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs, *Amendment to IEEE Std 802.11-2007*, pp. 1–126 (2008).
- [17] Sanchez, Isabel, M., Boukerche and Azzedine: On IEEE 802.11 K/R/V amendments: Do they have a real impact?, *IEEE Wireless Communications*, Vol. 23, No. 1, pp. 48–55 (2016).
- [18] McNair, J. and Zhu, F.: Vertical handoffs in fourth-generation multinet environments, *IEEE Wireless Communications*, Vol. 11, pp. 8–15 (2004).
- [19] Zhu, K., Niyato, D. and Wang, P.: Network Selection in Heterogeneous Wireless Networks: Evolution with Incomplete Information, *2010 IEEE Wireless Communication and Networking Conference*.
- [20] Khan, M. A., Toseef, U., Marx, S. and Goerg, C.: Game-Theory Based User Centric Network Selection with Media Independent Handover Services and Flow Management, *2010 8th Annual Communication Networks and Services Research Conference*.
- [21] Pervaiz, H.: A Multi-Criteria Decision Making (MCDM) network selection model providing enhanced QoS differentiation to customers, *2010 International Conference on Multimedia Computing and Information Technology (MCIT)*, IEEE, pp. 49–52 (2010).
- [22] Wang, S., Liu, H., Gomes, P. H. and Krishnamachari, B.: Deep Reinforcement Learning for Dynamic Multi-channel Access in Wireless Networks, *IEEE Transactions on Cognitive Communications and Networking*, Vol. 4, No. 2, pp. 257–265 (2018).
- [23] Hsu, A. C.-C., Wei, D. S. and Kuo, C.-C. J.: A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks, *IEEE wireless communications and networking conference*, pp. 105–110 (2007).
- [24] Li, M., Li, P., Huang, X., Fang, Y. and Glisic, S.: Energy consumption optimization for multihop cognitive cellular networks, *IEEE Transactions on Mobile Computing*, Vol. 14, No. 2, pp. 358–372 (2014).
- [25] Xiong, C., Li, G. Y., Zhang, S., Chen, Y. and Xu, S.: Energy-Efficient Resource Allocation in OFDMA Networks, *IEEE Transactions on Communications*, Vol. 60, No. 12, pp. 3767–3778 (2012).
- [26] Huang, H., Peng, Y., Yang, J., Xia, W. and Gui, G.: Fast beamforming design via deep learning, *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 1, pp. 1065–1069 (2019).
- [27] Wu, Q. and Zhang, R.: Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design, *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6 (2018).
- [28] Wu, Q. and Zhang, R.: Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming, *IEEE Transactions on Wireless Communications*, Vol. 18, No. 11, pp. 5394–5409 (2019).
- [29] Zhang, J., Yu, F., Wang, S., Huang, T., Liu, Z. and Liu, Y.: Load balancing in data center networks: A survey, *IEEE Communications Surveys and Tutorials*, Vol. 20, No. 3, pp. 2324–2352 (2018).
- [30] Katta, N., Hira, M., Ghag, A., Kim, C., Keslassy, I. and Rexford, J.: CLOVE: How I learned to stop worrying about the core and love the edge, *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, pp. 155–161 (2016).
- [31] Hart, P. E., Nilsson, N. J. and Raphael, B.: A formal basis for the heuristic determination of minimum cost paths, *IEEE transactions on Systems Science and Cybernetics*, Vol. 4, No. 2, pp. 100–107 (1968).
- [32] Shah, D., Sridhar, A., Bhorkar, A., Hirose, N. and Levine, S.: GNM: A General Navigation Model to Drive Any Robot, *arXiv*, (online), available from <https://arxiv.org/abs/2210.03370> (2022).
- [33] Yang, Y., Juntao, L. and Lingling, P.: Multi-robot path planning based on a deep reinforcement learning DQN algorithm, *CAAI Transactions on Intelligence Technology*, Vol. 5, No. 3, pp. 177–183 (2020).