

1X-05

運搬タスクにおけるロボット制御によるネットワーク性能向上手法の一検討

森 友紘<sup>†</sup> 新宮 裕章<sup>‡</sup> 藤橋 卓也<sup>‡</sup> 工藤 理一<sup>†‡</sup> 高橋 馨子<sup>†‡</sup> 村上 友規<sup>†‡</sup> 渡辺 尚<sup>‡</sup> 猿渡 俊介<sup>‡</sup>

<sup>†</sup>大阪大学工学部 <sup>‡</sup>大阪大学大学院情報科学研究科

<sup>†‡</sup>日本電信電話株式会社 NTT 未来ねっと研究所

1 はじめに

近年の5Gをはじめとして遅延を意識したワイヤレスネットワークの登場に相まって、センサだけでなく低遅延性が求められる自動運転車やロボットなどの移動体も無線を介してネットワークに接続し始めている。筆者らは、ロボットのタスク遂行効率とネットワーク性能の最適化を同時達成する仕組みの実現を目指している。本稿では、運搬タスクにおけるロボットの挙動制御に深層強化学習を用いることによって、高いネットワーク性能を保ちながらタスクを遂行する方法を提案する。

2 課題

これまでのネットワーク技術は、無線接続端末を有するユーザが自由に行動する中でユーザの利益を最大化するようにネットワーク側を最適化してきた。しかしながら、ヒトだけでなくロボットがネットワークを利用するようになった場合、ロボットとネットワークの協調が重要になると考えている。ロボットの挙動とネットワーク設定を同時に制御可能である場合、ロボットが与えられたタスクとネットワークの効率の両方を最適化することができる。

例として、図1に示すような複数台の無人搬送ロボットを用いた運送会社の倉庫での配達物の仕分けにおいて、ロボットの挙動やネットワークの設定を同時制御する場合を考える。各ロボットはすべて無線ネットワークに接続されて自身の位置情報やタスク情報を管理ノードと通信し、管理ノードはネットワークを通じてロボットを集中制御する。各ロボットに割り当てられるタスクはロボット間で分担しながら倉庫内のすべての配達物を適切な場所へと仕分けすることが挙げられる。このとき、各ロボットが効率的にタスクを遂行するためには、それぞれのロボットに対して積み下ろし位置までの最短経路を管理ノードが示せばよい。しかしながら、各ロボットと管理ノードとを接続する無線ネットワークの通信性能は各ロボットとアクセスポイントの距離、他のロボットの行動や障害物との位置関係の時間変化によって生じるフェージングやシャドウイングによって変化することが考えられる。通信品質の低下はロボットに適切なタイミングで指示が届けられないことにつながる。管理ノードが適切な行動と判断した動作指示が遅延や通信の切断により遅れた場合、ロボットが動作する時には適切な行動とならないことがある。余分な移動の増加により増加タスク効率が低下したり、障害物に衝突してしまうなどの安全面での問題につながることも考えられる。このような環境下では、各ロボットと管理ノードとの無線ネットワーク切断を防ぐ、スループットの低下を防ぐ無線ネットワーク側のチャンネル選択やロボットの行動を最適化することが有用となる。

3 運搬タスクのシミュレーション環境

本稿では2次元平面上を3台のロボットが通信しながら移動して荷物の仕分けを行う計算機シミュレータを実装した。荷物は食品系、衣服系、機械系、家具系のように種類があり種類ごとに格納場所が異なるものとした。図1は本稿で想定する環境である。想定した平面空間は横13m、縦7mで、ロボットの動ける範囲は図1の白色部分及び、A~Dの荷物の種類に応じた積みおろし地点、Loading Point(LP)で示した搬入口の範囲である。荷物はLPで配列を用いた管理を行うものとした。配列の初期状態は ABCDABCD... と規則的な並びで各5個ずつの20個の荷物とした。配列は3台のロボットで同じものを用い、あるロボットがLPに到達すると配列の先頭の荷物を受け取り、次にLPに到達したロボットはその時点で配列の先頭にある荷物を受け取るというように荷物の運搬順は固定とした。タスクの終了条件は3台のロボットで20個の荷物をそれぞれの指定地点に運び終えることとした。

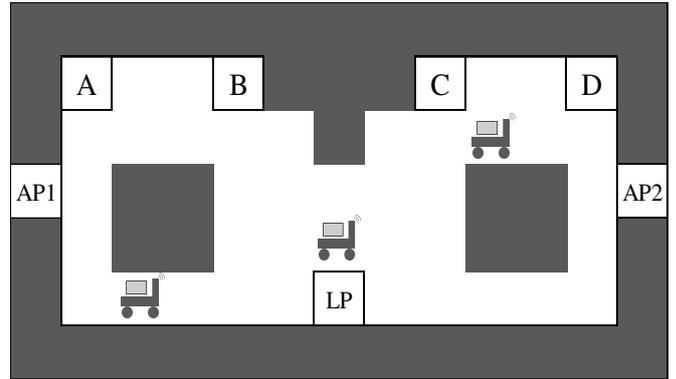


図1: 本稿で想定した環境

次に通信性能についての想定を示す。各ロボットは無線APに接続すると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。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。ここで、ロボット*i*が無線AP*m*に対して接続する場合、受信信号強度  $S_{i,m}$  [W] はフリスの伝達公式にしたがって次式で定まる。

$$S_{i,m} = P_t \left( \frac{\lambda}{4\pi d_{i,m}} \right)^2 \quad (1)$$

$P_t$  はロボット*i*の送信電力 [W],  $\lambda$  は送信電波の波長 [m],  $d_{i,m}$  はロボット*i*と無線AP*m*間の通信距離 [m]である。このとき、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] はシャノン容量にしたがって次のとおり定まる。

$$C_{i,m} = \frac{B}{L_m} \log_2 \left( 1 + \frac{S_{i,m}}{N} \right) \quad (2)$$

ここで、 $L_m$  は無線AP*m*に接続されるロボットの台数、 $B$  は帯域幅 [Hz],  $N$  はノイズフロア [W] である。

4 深層強化学習に基づくロボット挙動最適化

図4に提案するシステムの全体像を示す。提案手法で前提とするシステムはロボット部、ネットワーク部、ロボット・ネットワーク制御部により構成される。ロボット部はロボット・ネットワーク制御部から動作命令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。ロボット部は制御対象である作業車両やドローンなどのロボットそのものを指す。ロボット部は入力として操作、出力として状態情報を持つ。ロボット部の状態情報はロボットの位置、目標地点の位置などの情報を意味する。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して情報のやり取りを行う。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。ネットワーク部からの通信情報は Received Signal Strength Indicator (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

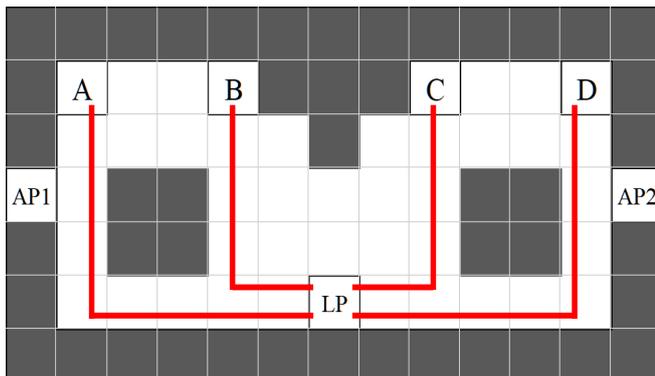


図2: 通信性能が高くなる運搬経路

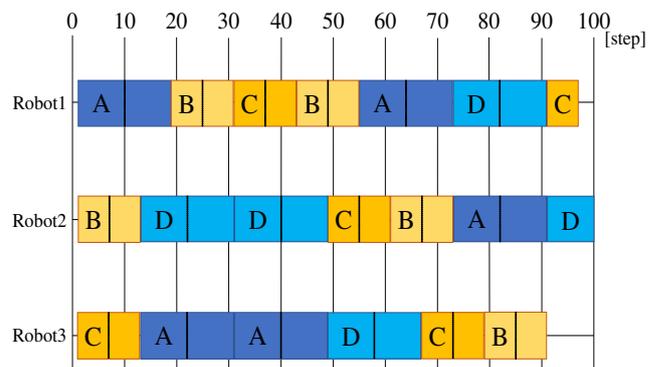


図3: 最高タスク効率時の運搬の様子

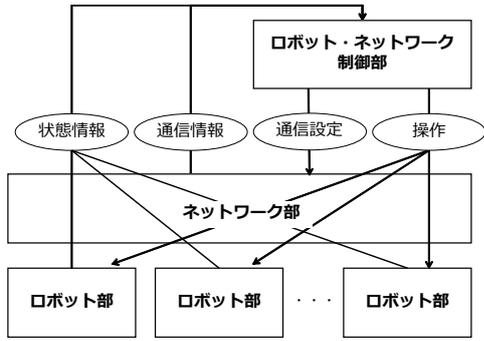


図 4: 提案手法の全体像

などの情報を意味する。ロボット・ネットワーク制御部は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操作や、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。

本手法では、深層強化学習を用いて適切なロボットの移動制御を推定する。本手法はロボットの自己位置とロボットの目的地を入力とした適切なロボットの移動制御の推定、推定したロボット操作の環境への反映、操作後のロボットのタスク遂行効率とネットワーク性能を元にした報酬の計算の3つの過程を繰り返して適切なロボットの移動制御を学習する。本手法では、ロボット  $i$  による時刻  $t$  の行動  $a_t^i$  を推定するために、ロボットの自己位置の縦軸座標と横軸座標、ロボットの目的地の縦軸座標と横軸座標を入力として与える。入力したロボットの自己位置と目的地に対して全結合層と活性化関数である ReLU (Rectified Linear Unit) を2回用いてロボットの自己位置と目的地に関する特徴を抽出する。ReLU は負の値を 0 とする活性化関数  $f(x) = \max(0, x)$  である。全結合層が出力した 1 次元ベクトルを半分に分割してそれぞれを状態価値関数・行動価値関数の学習に用いる。最後に、状態価値関数から得られた出力と行動価値関数の出力を結合した状態行動価値関数から適切なロボット  $i$  の行動  $a_t^i$  を推定する。行動  $a_t^i$  は上、下、右、左の4方向への1mの移動または停止の5種類とした。

学習モデルの実装には深層学習フレームワーク PyTorch を元にした深層強化学習ライブラリ PFRL を利用した。その他のパラメータは文献 [2] にしたがった。表 1 に学習に用いた報酬関数を示す。1 ステップごとの減点はタスクをなるべく少ないステップで終わらせるためのものである。侵入できない領域に入ろうとした場合は、ロボットの位置を更新しないものとした。ロボットが荷物を運搬していない状態で、搬入口のバッファに荷物が存在しない場合、いずれの行動をとった場合も報酬は 0 とした。

表 1: 報酬関数

条件	報酬
1 ステップの行動	-1
侵入不可領域への侵入の試行	-5
停止	-5
目的地への接近	+2
荷物の積み下ろし	+100
スループットが 0.2 Gbps を下回る	-3
目的地が搬入口かつ、荷物のバッファが空	上記を無視して 0

### 5 評価

報酬関数が学習結果に与える影響を評価するため、通信性能を考慮する「スループットが 0.2Gbps を下回ると減点」という項目の有無によりタスクの効率と通信性能にどのような影響を与えるかを検討した。

図 2 及び図 3 に本シミュレータの理想的な動きを示す。タスク効率としては少ない移動回数で荷物をすべて運ぶことが望ましい。図 3 に、

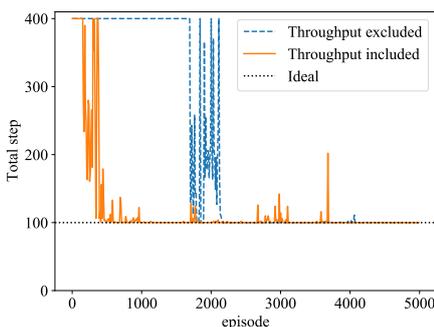


図 5: 総ステップ数

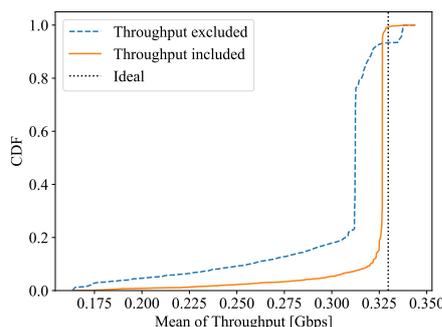


図 6: 平均スループットの CDF

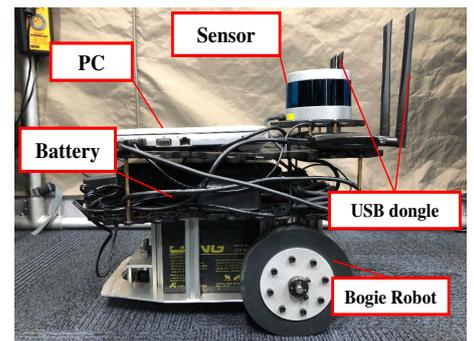


図 7: 実機の様子

本シミュレーションでタスク効率が最大となる荷物の運び方を示す。タスク開始から 1 ステップ目はプログラム上ロボットの移動後に積み下ろしを行うため停止の行動をとらせている。LP から A または D 地点へは 9 ステップの移動が最短であり往復には 18 ステップを要する。LP から B または C 地点へは 6 ステップの移動が最短であり、往復には 12 ステップを要する。運搬タスク開始から終了までに要するステップ数を総ステップ数とする、ロボットが配列順に受け取った荷物を上記のステップ数で運び続けることでタスク終了までの総ステップ数は最小となる。図 3 に示したようにロボットが 3 台の場合は 100 ステップでタスクが終了するため総ステップ数は 100 が理想値となる。図 2 に、各積み下ろし地点への経路の内、最も平均スループットが高くなる経路を示した。このシミュレータでの理想的な動作は図 2 の経路を用いて、図 3 に従い 3 台のロボットが荷物を運搬することである。

評価結果の図中に示す Throughput Excluded は報酬関数にスループットに関する項目を含まない提案手法、Throughput Included は報酬関数にスループットに関する項目を含む提案手法である。Ideal は理想的な挙動時の値である。

図 5 に、エピソード数に対するタスクが終了するまでに必要とした総ステップ数を示す。Throughput Excluded は 2300 エピソードで 100 ステップに収束した。Throughput Included は 1200 エピソードで 100 ステップに収束した。図 5 の結果より、以下の 3 つのことが分かった。1 つ目は、十分なエピソード数の学習を経た提案手法によってタスク効率の最適化がなされたことである。2 つ目は、報酬関数のスループットを考慮する項目の有無がタスク効率の収束値に影響を与えていないことである。3 つ目は、報酬関数のスループットを考慮する項目の有無が学習の収束に必要なエピソード数に影響を与えることである。報酬関数の項目が多いことで状態行動価値関数の更新が起りやすくなった結果、状態に対する最適な行動を絞り込みやすくなったことが考えられる。

図 6 に、5000 エピソードまでの各エピソードの平均スループットの累積分布関数を示す。Ideal の値は 0.330 Gbps であった。Throughput Excluded の 50%点は、0.327 Gbps であり、Ideal の 99%の性能であった。Throughput Included の 50%点は、0.312 Gbps であり、Ideal の 97%の性能であった。図 6 の結果より、以下の 2 つのことが分かった。1 つ目は、Throughput Included は Throughput Excluded と比較して、タスク効率を維持したままより高い通信性能を達成したことである。2 つ目は、報酬関数にスループットに関する項目を追加することで Ideal に漸近する通信性能を達成したことである。

### 6 おわりに

本稿では、深層強化学習を用いたロボットのタスク効率とネットワーク性能の最適化を行った。深層強化学習の報酬関数に通信性能に関する項目を追加することでタスク効率を維持したまま通信性能の向上ができたことを示した。現在、運搬する荷物の種類を選択するモデルの検討や、図 7 に示す実機による提案手法の評価を進めている。

### 謝辞

本研究は JSPS 科研費 (JP19H01101, JP20K19783), NTT 研究所の支援の下で行った。

### 参考文献

- [1] 新宮 裕章, 本生 崇人, 藤橋 卓也, 工藤 理一, 高橋 馨子, 村上 友規, 渡辺 尚, 猿渡 俊介, "ロボットの挙動制御によるネットワーク性能向上手法の検討", マルチメディア, 分散, 協調とモバイル (DICOMO2021) シンポジウム, pp.1-10, 2021.
- [2] Hessel, Matteo and Modayil, Joseph and Van Hasselt, Hado and Schaul, Tom and Ostrovski, Georg and Dabney, Will and Horgan, Dan and Piot, Bilal and Azar, Mohammad and Silver, David, "Rainbow: Combining improvements in deep reinforcement learning", Thirty-second AAAI conference on artificial intelligence, 2018.