

5T-09

ネットワークとロボットとの協調制御に関する初期的検討

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[‡]
[†]大阪大学工学部 [‡]大阪大学大学院情報科学研究科 ^{†‡}日本電信電話株式会社 NTT 未来ねっと研究所

1 はじめに

IoT (Internet of Things) の普及とともに、センサだけでなく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 [1]。筆者らは、ロボットの行動とネットワークの挙動を同時最適化する仕組みの実現を目指している。これに向けて、本稿では深層強化学習 DQN (Deep Q Network) [2] を用いてロボットの挙動を最適化する方法を提案する。

2 課題

図 1 (a) にヒトを中心としたこれまでのネットワーク技術を示す。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ーが自由に行動する中でユーザー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。図 1 (b) に、筆者らが目指している技術の最終ゴールを示す。ロボットの行動とネットワークの挙動を同時に操作可能である場合、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、多数の無人搬送車を用いて運送会社の配達物の仕分けにおいてロボットやネットワークの挙動を最適化する場合を考える。無人搬送車は全て無線ネットワークに接続されて集中制御される。この時、無線ネットワークから切断されると無人搬送車は挙動を停止する。また、他の無人搬送車の行動によってフェージングが変化したりシャドウィングが生じたりする。このような環境下では、無人搬送車の無線ネットワークが切断されないように無線ネットワーク側のチャンネルの選択や無人搬送車の行動を最適化することが有用となる。

本稿では、図 1 (b) に示したような最終ゴールの第一段階として、図 1 (c) に示すようなネットワークの状況に応じてロボットの行動の最適化が可能か検討する。図 2 に異なるチャンネルで接続可能な無線 AP (Access Point) が 2 台存在する環境でロボット 1 台が部屋を動き回りながら SLAM (Simultaneous Localization and Mapping) で部屋の空間データを更新し続ける例を示す。この単純な例では、ロボットは 2 つの無線 AP の RSSI (Received Signal Strength Indicator) を計測して RSSI の大きい方に接続するだけで収集するデータ量を最大化することができる。

しかしながら、上記のロボット 1 台の例を少し複雑化するだけで RSSI のみを用いた最適化が困難となる。例として、ロボットを 2 台にして空間の情報を 2 倍の速度で更新し続けることを考える。ロボットが 2 台存在する場合に、単純に RSSI が大きい方の無線 AP を用いると 2 つのロボットが同じ無線 AP に接続してチャンネルを共有することでスループットが減少する可能性が生じる。このような状況下で収集可能なデータ量を最大化するためには、相手のロボットの位置、相手のロボットが接続している無線 AP、各ロボットと各無線 AP 間の RSSI などの通信品質の情報が必要となる。さらに、現実的には外部のユーザーによる通信や周囲のユーザーの移動なども時々刻々と変化することを考えると、環境の変化に追従して最適化を行う仕組みが必要となる。

3 深層強化学習に基づくロボット通信選択手法

2 節の議論を基に、深層強化学習に基づくロボット通信選択手法の設計を行った。図 3 に提案手法の全体像を示す。提案システムは、ロボット情報収集部、通信情報収集部、ロボット制御部、ネットワーク、ロボットから構成される。ロボット情報収集部は、各ロボットからロボットの位置・向き・速度などの状態情報や RSSI の情報を収集する。通信情報



図 2: ロボットが 1 台の環境

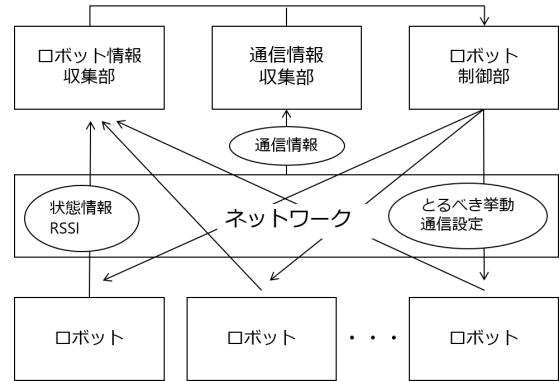


図 3: 提案手法の全体像

収集部は、無線 AP 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、遅延などの情報を収集する。ロボット制御部はロボット情報収集部と通信情報収集部で収集された情報を受け取って、ロボットに対して取るべき挙動や接続する無線 AP などの通信設定の指令を出す。

図 4 に本稿で想定する環境を示す。102 m 離れた 2 つの無線 AP 1, 無線 AP 2 間の直線上にある地点 A, 地点 B 間 100 m を 2 台のロボットが自律的に往復する。1 台目のロボットは地点 A, 2 台目のロボットは地点 A, 地点 B 間のランダムな点から移動を開始するものとする。1 台目のロボットは速度 10 m/s で地点 A, 地点 B 間を 30 秒間往復する。2 台目のロボットは 1 秒ごとに速度を 4~8 m/s の範囲で一樣に変化させながら地点 A, 地点 B 間を 30 秒間往復する。各ロボットは 1 秒ごとに接続する無線 AP を選択する。このとき、各無線 AP が使用する無線チャンネルは周波数が十分に離れているため、互いに干渉しないものとした。また、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 を選択して接続する。

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。

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

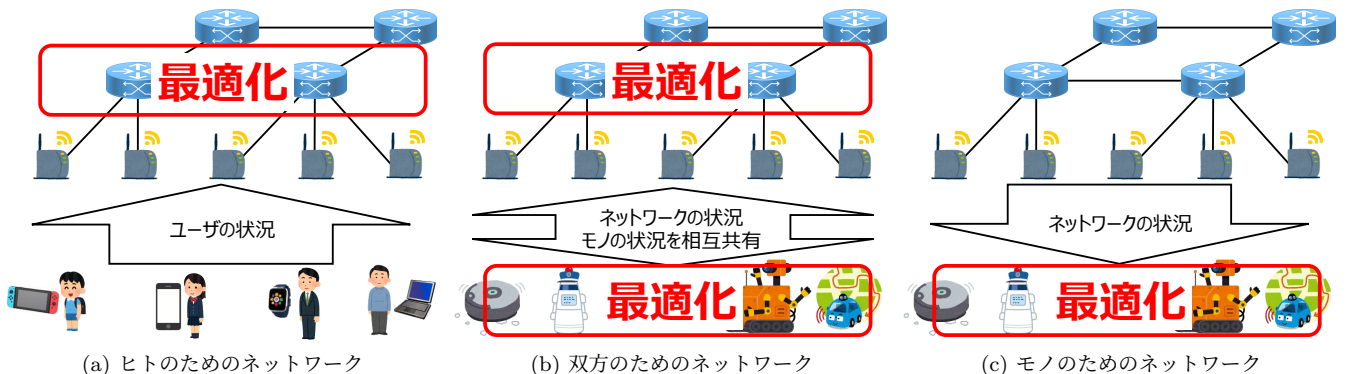


図 1: 既存のネットワーク技術と本研究で目指すネットワーク技術

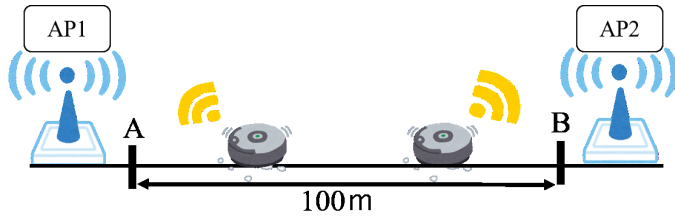


図 4: 本研究で想定した環境

P_i はロボット i の送信電力 [W], λ 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。

提案システムでは、無線チャンネルを介して 2 台のロボットから収集可能なデータ量を最大化、すなわち上り方向の通信スループットを最大化するために、深層強化学習を基にした適応的無線チャンネル選択手法を用いる。深層強化学習は各時刻 t における環境内でのロボットの状態 s_t , ロボットが方策 $\pi(a_t|s_t)$ にしたがって選択した行動 a_t , 選択した行動にもなつて得られた報酬 r_t を通じてロボットが最適な行動規則を獲得するための枠組みである。このとき、ロボットの目的は将来を通して累積報酬 $\sum_t \gamma^t r_t$ が最大となる方策を見出すこととなる。ここで、 γ は将来の報酬をどの程度割り引くかを表す割引率である。

DQN では環境内の状態 s_t においてロボットが行動 a_t を選択したときに将来的に得られる報酬の期待値をニューラルネットワークでモデル化した行動価値関数 $Q(s_t, a_t; \theta)$ によって定義する。適当な行動価値関数がニューラルネットワークから正しく推定できたと仮定すると $\arg \max_{a_t} Q(s_t, a_t)$ から各状態 s_t に対する最適な行動 a_t^* を選択できる。

このとき、ニューラルネットワークのパラメータ θ はベルマン方程式にしたがう下式の損失関数 $L(\theta)$ を最小化することで求めることができる。

$$L(\theta) = E \left[\left(r_{t+1} + \epsilon \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta') - Q(s_t, a_t; \theta) \right)^2 \right] \quad (3)$$

ここで、 ϵ は学習率、 θ' は更新前のパラメータである。

提案手法では、各時刻 t における各ロボットの位置を状態 s_t , 各位置でのロボットによる接続先無線 AP の選択を行動 a_t , 各ロボットの行動によって得られる合計通信スループットを報酬 r_t と定めた。また、行動価値関数を近似するニューラルネットワークとして複数の全結合層から構成される MLP (Multi-Layer Perceptron) を利用した。図 5 に提案手法で利用するニューラルネットワーク構造を示す。入力層は 2 ロボットの位置を示す 4 ノード、隠れ層は全結合層 3 層で各層は 64 ノード、出力層は各ロボットがいずれの無線 AP に接続する行動をとるべきかを示す 4 ノードとした。ここで、出力層のノード 1 は 1 台目のロボットが無線 AP 1 に接続・2 台目のロボットが無線 AP 1 に接続する行動、ノード 2 は 1 台目のロボットが無線 AP 1 に接続・2 台目のロボットが無線 AP 2 に接続する行動、ノード 3 は 1 台目のロボットが無線 AP 2 に接続・2 台目のロボットが無線 AP 1 に接続する行動、ノード 4 は 1 台目のロボットが無線 AP 2 に接続・2 台目のロボットが無線 AP 2 に接続する行動を表している。また、各全結合層の後に ReLU (Rectified Linear Unit) 活性化層および dropout (0.4) を導入した。

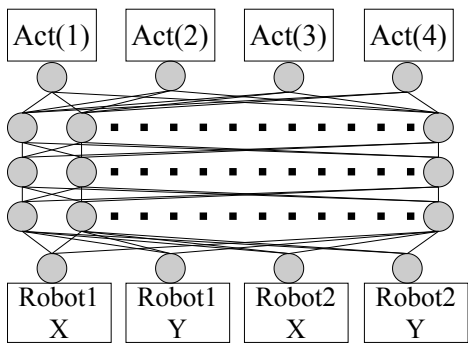


図 5: 本研究で利用するニューラルネットワーク構造

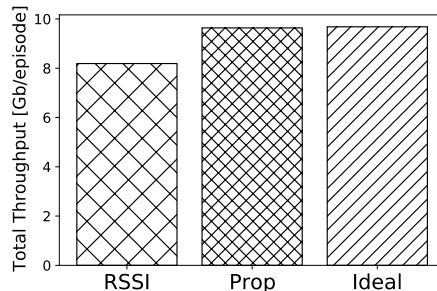


図 6: 各手法の合計スループット

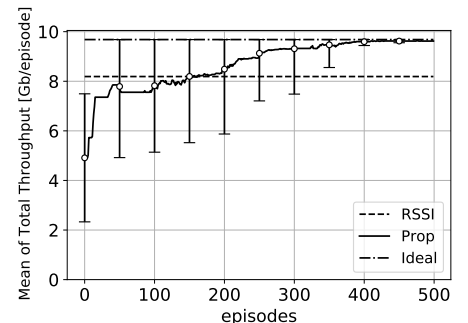


図 7: 各エピソードにおける合計スループット

提案手法の実装には深層学習フレームワーク PyTorch を用いて DQN を実装した。ハイパーパラメータとして Experience Replay に利用するバッチサイズを 30, RMSprop 最適化アルゴリズムに利用する学習率を 0.00015, 平滑化定数 α を 0.95 と定めた。その他のハイパーパラメータは PyTorch のデフォルトパラメータにしたがった。また、学習中に局所的な行動に陥らないようにするため、 ϵ -Greedy 法に基づいてランダムに行動を選択する。

4 性能評価

提案方式による効果を議論するために、トレーニングデータを用いて学習済モデルを生成するとともにテストデータからスループット性能を評価した。トレーニングデータとして各ロボットによる 30 秒間の行動を 1 エピソードとする 500 エピソードを与えた。行動価値関数は 20 エピソードごとに更新するものとした。テストデータにおいては、1 台目のロボットが地点 A, 2 台目のロボットが地点 A から 1 m 離れた地点から移動を開始するものとした。1 台目のロボットは速度 10 m/s で地点 A, 地点 B 間を 30 秒間往復する一方で、2 台目のロボットは 1 秒ごとに速度 4 m/s で地点 A と地点 A, B 間の中点を 30 秒間往復するものとした。各ロボットが無線通信に利用する送信電力を 0.001 W, 帯域幅を 22 MHz, ノイズフロアを $1.0 \cdot 10^{-12}$ W とした。

また、提案手法を相対的に評価するための比較手法として、各ロボットが無線 AP から受信した信号強度 RSSI に基づいて接続する無線 AP を選択する手法と、合計スループットが最大化する理想的な無線 AP 選択手法を用意した。

4.1 最適性

図 6 に各比較手法によって得られる上り方向の合計スループットを示す。無線 AP から得られた RSSI に基づいて各ロボットが接続先の無線 AP を選択する手法を RSSI, 提案手法を Prop, 理想的な無線 AP 選択手法を Ideal とした。ここで、提案手法はすでに 500 エピソード分のトレーニングデータから学習済モデルを得られているものとした。評価結果から提案手法は 500 エピソード分のトレーニングデータから適切な行動価値関数を推定することで、通信干渉が発生する環境下においても理想的な無線 AP 選択手法に漸近する性能を達成したことが分かる。

4.2 収束するまでの時間

図 7 に学習済エピソード数に対する上り方向の合計スループットを示す。ここで、合計スループットは 10 回の試行平均である。また、エラーバーは、95%信頼区間を示している。評価結果から以下の 2 つのことが分かる。1 つ目は 350 エピソードを学習した時点で RSSI にしたがって接続先無線 AP を選択する手法と比較して提案手法が優れた性能を達成できていることである。2 つ目は 400 エピソードを学習した時点で提案手法が理想的な無線 AP 選択手法に漸近する性能を達成できていることである。

5 おわりに

本稿では、深層強化学習を用いたロボットのネットワーク状況に基づく適応的チャンネル選択方式を提案し、その効果を検証した。今現在、実験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, JP18H03231, JP20K19783) の支援の下で行った。

参考文献

- [1] 工藤 理一, コ克蘭 マシュー, 高橋 馨子, 井上 武, 水野晃平, “移動ロボットによる無線 LAN 通信品質予測”, 電子情報通信学会センサネットワークとモバイルインテリジェンス研究会, 2020.
- [2] V. Francois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, “An Introduction to Deep Reinforcement Learning,” arXiv 1811.12560, 2018.