

4-4-3 強化学習を用いた自律型モビリティ利活用データ集配新技術の研究

4-4-3 *Reinforcement Learning based Data Query Scheme for Contents Synchronization between Mobilities*

長谷川 聡 荘司洋三

HASEGAWA So and SHOJI Yoza

ミリ波通信において、移動するモビリティ同士——例えば車やロボットがすれ違いざまにデータを送受信する時間は極めて限られる。扱うデータをモビリティ間で事前に決めておくことで、通信時間を最大限活用した効率のよいデータ転送が可能となる。本稿ではモビリティ間で事前にデータの照会を行い、差分及び更新チェックを行うことを Multi-Armed Bandit (MAB) 問題としてモデル化し、強化学習アルゴリズムである Tug-of-War (TOW) を用いて解くデータ照会手法を紹介する。シミュレーションによる性能検証結果から、セグメント状に管理された 10 のデータにおいて、データ別に更新頻度を設定し、そのばらつきが大きい場合、強化学習を用いたデータ照会による更新データの検出率はランダムなデータ照会と比較して最大 40 % 高いことが示された。本手法はデータの更新頻度や重要度の学習が可能であることから、ミリ波による効率的なデータ更新・同期が期待できる。

Mobilities, such as vehicles or robots, can transmit and receive data using millimeter-wave communications only for a limited period of time. By predetermining the data to be handled between mobilities, the efficient data transfer is possible, and this time can be utilized maximally. This paper introduces the modeling of Multi-Armed Bandit (MAB) problem of difference and update check by prior data query between mobilities and the data query technique using Tug-of-War (TOW), a reinforcement learning algorithm. From the result of performance verification by simulation, it is shown that the detection rate of the update data by the data query using reinforcement learning is up to 40% higher than using random query, when the update frequency was set for every data in 10 data managed in the segment state and the dispersion is large. The efficient data update and synchronization by the millimeter wave can be expected, because this scheme can learn the update frequency and the importance of the data.

1 まえがき

大規模に分散した大容量データ——例えば高解像度カメラで長時間撮影した動画データやセンサで取得した膨大なログデータを携帯回線を用いた従来の方法で収集したり、複数ユーザで共有するには多大な時間を要する。こうした問題を解決するため、筆者らはミリ波通信と Store-Carry-Forwarding (SCF) 技術により物流網にデータの流通を託す Piggy-back Network を提唱している。先行研究[1][2]では実際に都市部を走行する車両の走行データを分析し、携帯回線 (4G) や IEEE 802.11.ac を利用するよりも高速なエンドツーエンドでのデータ転送が可能であることを主張している。

たとえば、100 Gbytes のデータを受け取って 2 km 先の宛先に届けるとき、20 Gbps のミリ波通信でデータを受け取り、宛先の元まで 20 km/h で移動すれば、実質 2 Gbps のエンドツーエンドのスループットが達成される。近年、ロボットの社会実装が盛んになってきており、同技術はロボット同士の小さなコミュニティにも適用が期待できる。

ロボットは様々な高精度センサを搭載し、周辺の情報を取得したり、我々消費者向けのニュースや動画といった膨大なコンテンツを提供するなど頻繁に大容量データを扱う。このようなロボット同士が互いにデータを共有・分析することで、気づきが生まれ、より高効率な稼働や、より高品質なサービスの提供が可能

になる。ここで問題となるのが、データ共有に必要となるミリ波の通信性とロボットの移動である。ミリ波通信は高速かつ大容量のデータ通信が可能である一方で、通信距離が短く、ロボット同士が相互に移動する状況では通信時間が極めて限られる。そのため限られた通信時間を最大限活用するためには、事前に送受信の対象となるデータを決めておくことが適切である。

本稿では自律移動ロボット同士によるミリ波 SCF データ転送ネットワークのための転送データの事前交渉技術について紹介する。具体的にはセグメント状に管理されるデータを同期するためのデータの差分及び更新チェックの Multi-Armed Bandit (MAB) 問題 [3] としてのモデル化、MAB アルゴリズムである Tug-of-War (TOW) ダイナミクス[4]-[8] ベースのデータ紹介手法を紹介する。

2 システムモデル

図1は複数の自律移動するロボットで構成されたネットワークを示している。ロボットはマイクロ波やミリ波といった複数の通信方式での通信が可能であり、それぞれを用途によって使い分ける。ロボットの稼働

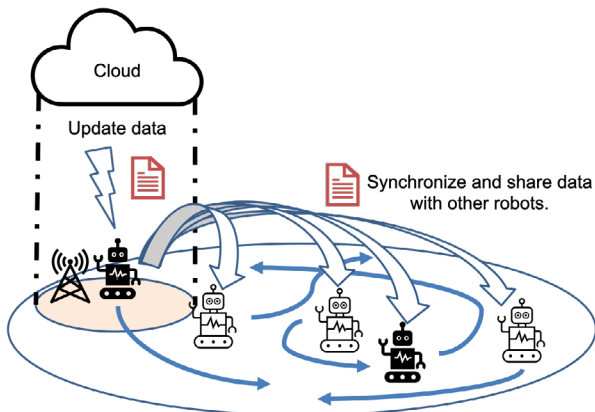


図1 自律移動型ロボット間のデータ同期・共有のためのネットワーク

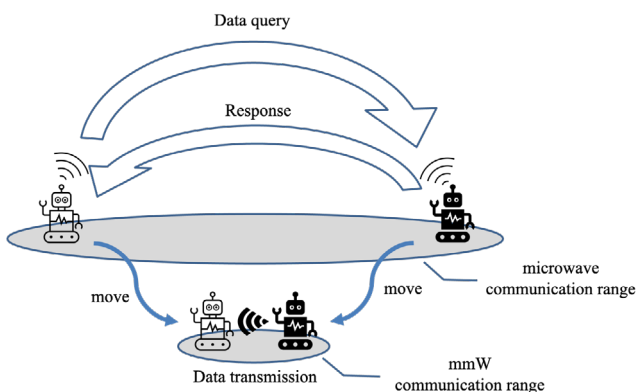


図2 2台のロボットによるデータ照会からデータ送信までの一連の流れ

エリア内には外部との通信を可能にする基地局が設置されてもよく、一部のエリアにおいてはクラウドやサーバとの通信が可能である。例えば、特定エリアに入ったロボットがクラウド等から更新データを受信したときや、あるロボットが自身で更新した内部データをネットワーク全体でこのデータを同期するのが目的となる。本システムにおいてはLTEや既存のマイクロ波通信では送受信に時間を要するサイズの大きな更新データを扱うことを想定する。

図2は2台のロボット間でデータを転送し、同期する一連の動作を表している。ロボットが他方のロボットのマイクロ波通信域に入ると、データを共有してもらいたいロボットはもう一方の提供元ロボットに特定のデータの照会を行う。この際、容量の大きな元データそのものを送信するのではなく、元データの内容を表す小さなメタ情報をやりとりする。照会を行ったロボットは自身の保持データと比較することで、今後更新すべきデータであるかどうかを判断する。続いてロボット同士が近接してミリ波通信域に入った場合、先ほどの照会情報に基づいてデータの転送を行う。

データ転送はロボット同士の近接を前提としているため、実際ロボット同士が近接できなかった場合にはこれらの事前準備は徒労に終わる。そこで、[9]の研究のような移動予測技術を用いることで、近接可能なロボットと交渉を行うことが有効的と考えられる。

図3はロボット間のデータ照会の様子である。ロボットは内部にデータベースを保持しており、 K 個のセグメントによって構成されていることを想定する。

ロボットは一度に1セグメントのデータコンテンツについて問い合わせることができ、もう一方のロボットは問合せを受けたデータのメタ情報を返す。この図では、ロボットがセグメント3のデータを問い合わせ、もう一方のロボットがそれに対して応答している様子を示している。問合せを行ったロボットはこの応答結

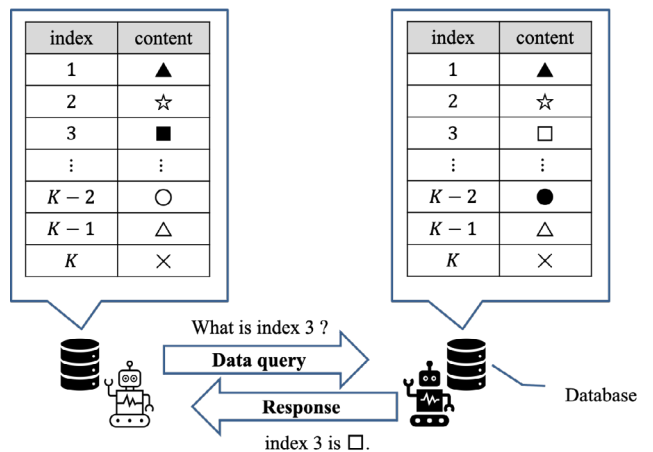


図3 2台のモビリティ間のデータ照会スキーム

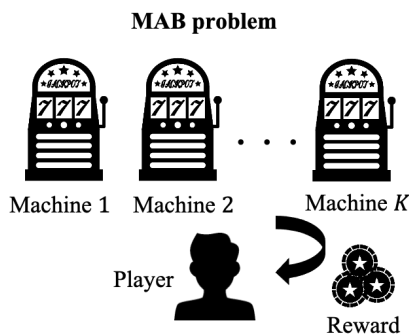
果から自身の持つデータと照らし合わせることで相手が保持するデータ内容と一致しているかを判断する。

ロボットはクラウドからのデータ配信やセンサで周囲の環境データを取得することで、データ更新が発生する可能性がある。これはすなわちロボットが保持するデータがある確率で変更され、セグメントごとに更新頻度に偏りが生じる可能性があるということであり、ネットワーク内のデータを同期し最新化するためには照会を行うロボットはこの更新されたデータを見つける必要がある。このデータ照会モデルは MAB 問題と一致する。

3 MAB アルゴリズム

MAB 問題は自身が所持するコインを消費しながら複数のスロットマシンをプレイし（アームを引く）、最大限の報酬（コイン）を得る問題である。マシンはそれぞれ固定の報酬確率に従って報酬を出し、プレイヤーはその事前情報を知らない。プレイヤーは各マシンをプレイするごとに、マシンの傾向すなわち報酬確率を把握していく。報酬確率の高いマシンを“探索”することとマシンをプレイして報酬を得る“活用”のバランスを決めるアルゴリズムが MAB アルゴリズムである。

図 4 は MAB 問題とデータ照会の対応を示しており、前章で説明したデータ照会モデルは、問合せを行うロボットが相手ロボットからの照会結果を基に、更新すべきデータを探索するという MAB 問題となっている。ここでデータ照会モデルが MAB 問題と異なる点は、確率的に行われたデータの更新は次回のデータ照会にも引き継がれるということである。次に MAB アルゴリズムである TOW ダイナミクスについて説明する。



3.1 TOW ダイナミクス

TOW は式 (1) で表される推定報酬 Q_k に従った強化学習型のアルゴリズムである。

$$Q_k(t) = N_k(t) - (1 + \omega)L_k(t) \tag{1}$$

ここで $N_k(t)$ はラウンド t までのマシン k の累計プレイ回数、 $L_k(t)$ はラウンド t までのマシン k の累計プレイ回数のうち報酬が出なかった回数、 ω は重みパラメータを示している。また、式 (1) は次のように書き換えられる。

$$Q_k(t) = Q_k(t - 1) + \Delta Q_k(t) \tag{2}$$

$$\Delta Q_k(t) = \begin{cases} +1 & (\text{報酬あり}) \\ -\omega & (\text{報酬なし}) \end{cases} \tag{3}$$

マシン k の推定報酬確率は以下で表される。

$$p_k(t) = \frac{R_k(t)}{N_k(t)} \tag{4}$$

ここで $R_k(t) = N_k(t) - L_k(t)$ であり、報酬を受け取った回数を表している。先ほどの重みパラメータ ω は以下のように設定することが最適であることが分かっている。

$$\omega = \frac{p_{1st}(t) + p_{2nd}(t)}{2 - (p_{1st}(t) + p_{2nd}(t))} \tag{5}$$

TOW の意思決定は変位 $X_k(t)$ によって行われ、推定報酬 $Q_k(t)$ を用いて以下で表す。

$$X_k(t) = Q_k(t - 1) - \frac{1}{K - 1} \sum_{k' \neq k}^K Q_{k'}(t) + \text{osc.} \tag{6}$$

ここで osc. は重みパラメータである。次回のラウンド $t + 1$ では、 $k^* = \arg \max_{k \in K} X_k(t + 1)$ となるマシン k^* を選択することで、探索と活用の意思決定が行われる。

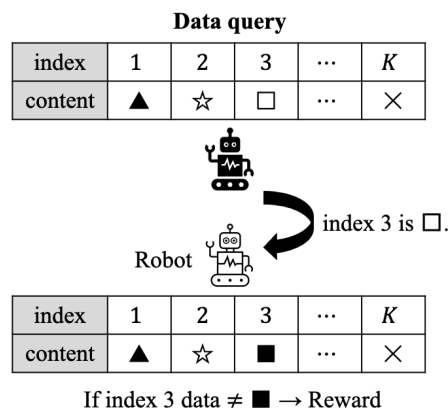


図 4 MAB 問題とデータ照会の比較

4 NICT 総合テストベッドの新たな可能性に向けた研究開発

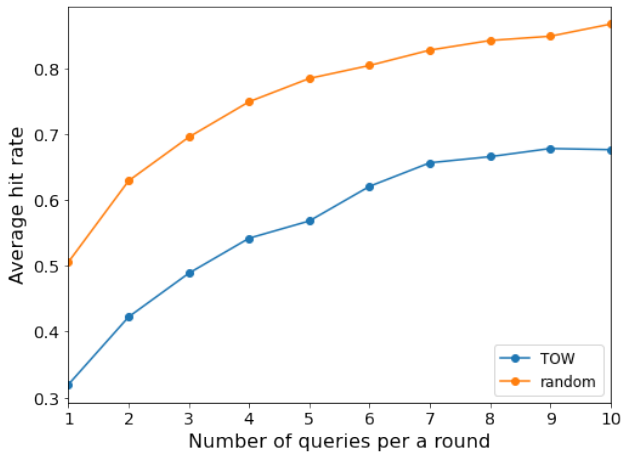


図5 1ラウンドあたりの照会回数と更新データの検出率（データ更新確率が一樣である場合）

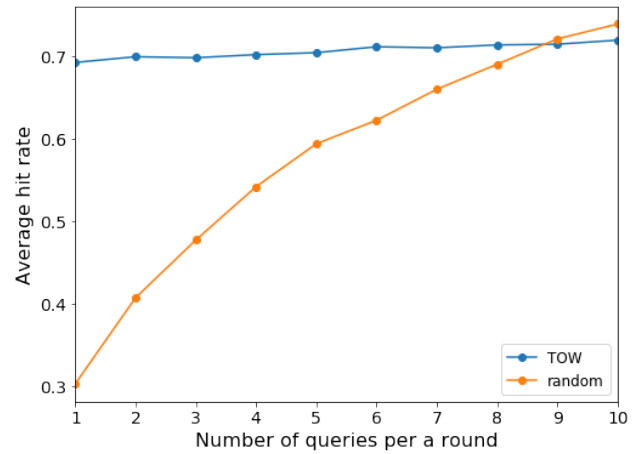


図7 1ラウンドあたりの照会回数と更新データの検出率（データ更新確率が1データと他データで大きな偏りがある場合）

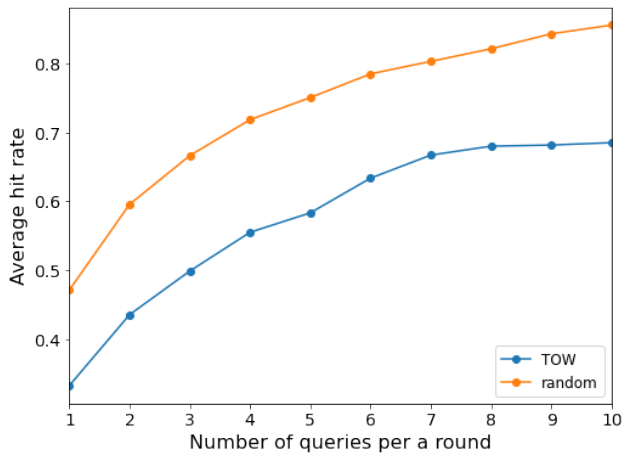


図6 1ラウンドあたりの照会回数と更新データの検出率（データ更新確率が階段状である場合）

つけたデータ更新確率における結果である。こちらもTOWは誤ったデータの更新率を学習したことにより、ランダムに照会した場合における結果の方が高くなっている。

最後に、図7は一つのセグメントのデータの更新頻度が高く、その他のセグメントのデータの更新率が一定のシナリオである。ロボットの稼働する場所や時間帯によって、こうしたデータの更新頻度に偏りが生じることは大いに想定される。この結果によると、照会回数が少ない場合のTOWの平均ヒット率が高く、照会回数に関わらず一定の平均ヒット率を示している。このことからデータの照会数に限りがあり、データの更新率に大きな偏りを持つ場合、TOWを用いたデータ照会スキームが有効であることが判明した。

4 性能評価

各ラウンドタイムに対して1セグメントのデータが更新されていくようなシステムを想定し、各セグメントに対して更新頻度が異なる3つのシナリオ別にデータ照会シミュレーションを行った。まず、図5は各データセグメントに対して、更新率が等しいシナリオにおける結果である。各タイムラウンドにおけるデータ照会の回数に対する平均ヒット率——照会データが更新すべきデータであったことを示している。この結果から、データの照会回数が増えることがヒット率の増加につながることを示している。TOWとランダムの照会スキームを比較するとランダム手法の場合による平均ヒット率が約20%高いが、これはデータの更新が一樣に発生する場合だと、特定のデータが更新されやすいと判断し、誤った学習が進んでしまうためであると考えられる。

続いて、図6はセグメントに順位付けされた重みを

5 むすび

本稿では、来るBeyond 5G時代でのミリ波SCFネットワークにおけるデータ同期のためのモビリティ間事前交渉技術として、機械学習型MABアルゴリズムであるTOWを用いたデータ照会スキームを紹介した。

シミュレーションより、10セグメントで構成されデータの更新頻度に大きな偏りがあり、かつ、データの照会回数に限りがある場合、本手法はランダムなデータ照会手法よりも最大約40%の更新データの検出が行えることを示した。様々なデータの更新頻度に対応できるようなアルゴリズムの改良が今後の課題である。

【参考文献】

- 1 Y. Shoji, W. Liu, and Y. Watanabe, "Community-based "Piggy-back Network"utilizing Local Fixed & Mobile Resources supported by Heterogeneous Wireless & AI-based Mobility Prediction," Proc. IEEE 91st Vehicular Technology Conference (VTC2020-Spring), pp.1-5, 2020.
- 2 Y. Watanabe, W. Liu, and Y. Shoji, "A Demonstrative Study on the Potential of Store-Carry-Forward-Based Contents Delivery by a Beverage Supplier's Logistics Network," Proc. International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp.65-70, 2019.
- 3 H. Robbins, "Some aspects of the sequential design of experiments," Bulletin of the American Mathematical Society, vol.58, pp.527-535, 1952.
- 4 S.-J. Kim, M. Aono, and E. Nameda, "Efficient decision-making by volume-conserving physical object," New Journal of Physics, vol.17, 083023, 2015.
- 5 S.-J. Kim, M. Aono, and M. Hara, "Tug-of-war model for the two-bandit problem: Nonlocally-correlated parallel exploration via resource conservation," BioSystems, vol.101, pp.29-36, 2010.
- 6 S.-J. Kim, T. Tsuruoka, T. Hasegawa, K. Terabe, and M. Aono, "Decision maker based on atomic switches," AIMS Materials Science, vol.3, pp.245-259, 2016.
- 7 S.-J. Kim, M. Naruse, and M. Aono, "Harnessing the Computational Power of Fluids for Optimization of Collective Decision Making," Philosophies, vol.1, pp.245-260, 2016.
- 8 S.-J. Kim and M. Aono, "Amoeba-inspired algorithm for cognitive medium access," NOLTA, vol.5, pp.198-209, 2014.
- 9 W. Liu and Y. Shoji, "Edge-Assisted Vehicle Mobility Prediction to Support V2X Communications," IEEE Transactions on Vehicular Technology, vol.68, no.10, pp.10227-10238, 2019.



長谷川 聡 (はせがわ そう)

総合テストベッド研究開発推進センター
 ソーシャル ICT システム研究室
 有期研究技術員
 IoT、機械学習、ネットワーク



庄司洋三 (しょうじ ようぞう)

総合テストベッド研究開発推進センター
 ソーシャル ICT システム研究室
 室長
 博士(工学)
 ミリ波通信システム、光通信システム