

LoRaWANにおける深層強化学習を用いた直交リソース割り当て法における報酬値の影響に関する検討

Impact of Reward Function on DQN-based Orthogonal Resource Allocation in LoRaWAN

相原 直紀¹ 安達 宏一¹ 田久 修² 太田 真衣³ 藤井 威生¹
Naoki Aihara Koichi Adachi Osamu Takyu Mai Ohta Takeo Fujii

電気通信大学¹ The University of Electro-Communications 信州大学² Shinshu University
福岡大学³ Fukuoka University

1 まえがき

多数の無線端末が自立分散的に周波数資源を共用する環境において、端末間干渉の軽減は重要である。本稿では筆者らが以前提案した、CSMA/CAを適用したLoRaWANシステムにおける深層強化学習を用いたリソース割り当て手法[1]において、より優れた特性を得るための新たな報酬関数を検討し、その特性をより現実的なトラフィックモデル環境下で評価する。

2 システムモデル

本稿では、固定配置された1個の情報集約局(FC)へ、 N 台の端末(集合 \mathcal{N})がパケットを送信する上りリンクを想定する。LoRaWANシステム全体で K 個の直交周波数チャネル(集合 \mathcal{K})を利用できるものとする。各端末は文献[2]に従いセンシングによる定期的パケットとイベント観測によるランダムパケットの2種類を生起する。パケット衝突が発生した場合は、受信信号対雑音電力比(SNR)と各拡散係数(SF)における干渉軽減を考慮した受信信号対干渉電力比(SIR)がそれぞれ閾値[3]を超えていれば受信成功となる。

本提案では、FCもしくはこれにつながるサーバで深層強化学習を行い、以下で与えられるQ報酬を最大化するようリソース割り当ての組み合わせを探索する。無線端末 $n \in \mathcal{N}$ に対するQ学習に必要な状態集合 \mathcal{S} 、行動集合 \mathcal{A}_n 、Q報酬値 $Q_{n,k}$ を以下のように定義する。

状態 \mathcal{S} エポック t において各端末に割り当てられている周波数インデックスの組み合わせ

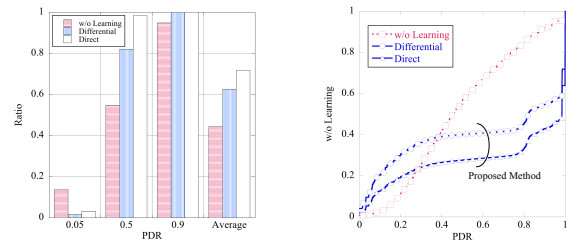
行動 \mathcal{A}_n 次エポック $t+1$ において端末 n に割り当てられる周波数インデックス $k_n \in \mathcal{K}$

Q報酬値 Q_{n,k_n} 他端末を考慮した受信成功パケット数であり、以下の式で与えられる。受信成功パケット数の差分を報酬とする $Q_{n,k_n,diff}$ と受信成功パケット数をそのまま報酬とする $Q_{n,k_n,dir}$ の2種類を検討する。

$$\begin{cases} Q_{n,k_n,diff} &= (D_{n,t+1} - D_{n,t}) + \\ & \nu \times \frac{\sum_{n' \in \mathcal{N} \setminus n} (D_{n',t+1} - D_{n',t})}{N-1} \\ Q_{n,k_n,dir} &= D_{n,t+1} + \nu \times \frac{\sum_{n' \in \mathcal{N} \setminus n} D_{n',t+1}}{N-1} \end{cases} \quad (1)$$

ここで $D_{n,t}$ はエポック t で端末 n から受信したパケット数、 ν は利己係数であり、次式で与えられる。

$$\nu = \tanh \left(\frac{D_{n,t+1}}{\min_{n' \in \mathcal{N} \setminus n} D_{n',t+1}} \right). \quad (2)$$



(a) 各 PDR 値の割合 (b) PDR の CDF

図 1: PDR 特性

FCはQ学習の結果から、 ϵ -greedy法を用いて各端末に対し周波数リソースを割り当てる。

3 計算機シミュレーション

計算機シミュレーションでは、 $3 \times 3 \text{ km}^2$ の領域内に $N = 100$ 個の端末をランダムに配置し、 $K = 8$ 個の周波数リソースを用いて通信するものとする。各端末に割り当てられるSFは、受信可能なSFの中で最も高いビットレートとなるものを選択する。通信路はパスロスと空間相関を有するシャドウイングに従うものとする。比較のために、各端末がランダムに送信周波数チャネルをホッピングする場合の特性も示す。図から分かるように、提案法を用いることによって、ランダム割当てと比較して平均PDRを28%程度向上出来ていることが分かる。PDRが0に近い領域ではランダム割当ての特性が良好となるが、これはランダムホッピングによりチャネルが劣悪なノードも確率的に干渉を回避できることが原因である。また、報酬関数としては差分を用いずに受信成功パケット数を直接学習する $Q_{n,k_n,dir}$ の特性が良好になった。これは各UEに対して差分情報を用いることによるQ報酬の変動を抑制することができるためである。

4 まとめ

本稿では、DQNに基づくリソース割り当て手法におけるQ報酬値についての検討を行い、現実的なトラフィックモデルでの評価を行った。シミュレーション結果から、提案手法はランダムな割当てと比較して28%程度平均PDRを向上できることを確認した。

謝辞本研究開発は総務省SCOPE(受付番号175104004)の委託によるものである。

参考文献

- [1] 相原直紀, 他, 信学技報, vol. 118, No.435, RCS2019-261, pp.109-114
- [2] V. Gupta, et. al, in Proc. IEEE Globecom, Singapore, 2017, pp. 1-6
- [3] C. Goursaud, and J.M. Gorce, EAI Endorsed Trans. Internet Things, vol. 1, no. 1, pp. 1-11, Oct. 2015.