6U-08

強化学習を用いたモバイルデータオフローディング手法自律獲得の検討

望月大輔† 町田樹‡ 峰野博史†

静岡大学情報学部† 静岡大学大学院総合科学技術研究科‡

1. はじめに

Internet of Things の普及や携帯端末の性能向上によるコンテンツの多様化に伴い，モバイルデータ通信需要は増加傾向にある．通信キャリアは通信インフラの負荷を分散するため，Wi-Fi スポットを設置しモバイルデータオフローディングに取り組んでいる．一方，モバイルデータ通信は時間帯や地域で通信インフラへの負荷に偏りがあり，モバイルデータ通信における帯域利用効率（以下，帯域利用効率）が低下する課題がある．帯域利用効率を最大化する方法として，携帯電話基地局(eNB：evolved Node B)負荷を分散するために携帯端末(UE：User Equipment)の送信レートを制御する必要がある．この背景の下，モバイルデータ通信の遅延耐性に着目し，UE の送信レートを制御させ，帯域利用効率の向上を目的とした Mobile Data Offloading Protocol (MDOP)[1]が提案されている．MDOP は時間的，空間的，通信路的の三つの方法で eNB 負荷を分散させるプロトコルである．しかし， MDOP の時間的オフローディングにおける環境に適した送信レート制御手法は確立されていない．

本稿では，MDOP の時間的オフローディングにおける帯域利用効率を最大化するため，強化学習を用いた送信レート自律獲得制御手法を提案する．UE が環境に適した送信レートを学習することで，適切な送信量や送信タイミングを自律制御可能とし，帯域利用効率の向上を図る．

2. 関連研究

帯域利用効率を高める送信レート制御の先行研究として，Quality of Service(QoS)制御で通信インフラへの負荷を分散させる User Plane Congestion Management [2]を模擬した手法[3]やビデオデータなどの短い遅延を許容するデータを要求する各 UE に対して等しい帯域を割り当てる手法[4]が提案されている．しかし，[3]は QoS の状態，[4]は多様なコンテンツが混在する環境下で帯域利用効率に偏りがある．また，コンテンツの遅延耐性を考慮していないため，送信レート制御に向上の余地がある．

モバイルデータ通信に強化学習を適用した先行研究として，強化学習手法の一つである Q 学習[5]を，マクロセルとフェムトセルが混在するヘテロジニアスネットワークのチャネル選択に適用した手法[6]が挙げられる．評価の結果から，チャネル選択を機械的に行う手法に，強化学習を適用することで，フェムトセル間の干渉を緩和することを示している．

[3][4][6]から，既存研究では考慮されていなかったコンテンツの遅延耐性に着目し，機械的に帯域を割り当てる送信レート制御に強化学習を適用することで，更なる帯域利用効率の向上が期待できる．

Mobile Data Offloading Protocol using reinforcement leaning
†Daisuke Mochizuki,‡Tatsuki Machida, †Hiroshi Mineno
†Faculty of Informatics, Shizuoka University
‡Graduate School of Integrated Science and Technology, Shizuoka University

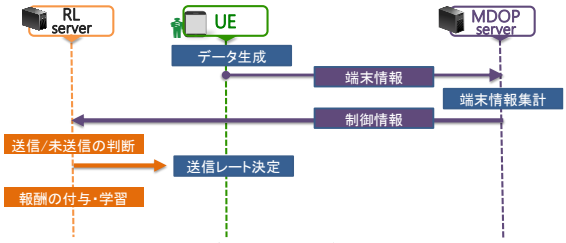

図 1：提案手法の通信フロー

3. 提案手法

3.1 概要

本稿では，MDOP の時間的オフローディングにおける帯域利用効率を向上するため，強化学習を用いた送信レート自律獲得制御手法を提案する．**図 1** に提案手法のアップロード時の通信フローを示す．MDOP の時間的オフローディングはトラフィックが特定の時間帯に偏る特性に着目して UE の送信レートを制御し eNB への負荷を分散させる．UE が遅延耐性を持つアプリケーションを一時的に蓄積し，通信インフラの状況に応じて UE と MDOP サーバが送受信する制御情報を元に送信レート制御することで，モバイルデータオフローディングを行う．端末情報には UE が保持しているコンテンツ量や接続先 eNB 情報，制御情報には eNB の負荷情報などの送信レートを決定するために必要な情報が含まれる．

3.2 送信レート制御手法

提案手法では，試行錯誤しながら行動を最適化する強化学習手法の一つである Q 学習を用いる．Q 学習は，ある状態sにおいて取りうる行動aの価値を行動価値関数$Q(s,a)$として定量化し，試行錯誤しながら Q 値を最大化するように逐次更新することで，各状態における行動を最適化する．式(1)に Q 学習における Q 値の更新式を示す．

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right) \quad (1)$$

式(1)のrは状態sで選択した行動aの良さを表す報酬であり，αは学習率$(0 \le \alpha \le 1)$，γは割引率$(0 \le \gamma \le 1)$とよばれる学習時に用いるパラメータである．また，学習の過程において，常に最善の行動を選択すると，局所解に陥る可能性がある．そのため，行動の選択に ε -greedy 法を採用することで，局所解に陥ることを回避する．

提案手法では，図 1 中の Reinforcement Learning Server(RL サーバ)が式(1)で送信レートを決定し，UE は RL サーバによって通知された送信レートを元にデータを送信する．RL サーバは eNB や UE の状況を状態とし，状態から帯域利用効率を最大化する送信レートを決定する．しかし，適切な送信レートが未知の場合，RL サーバが帯域利用効率を最大化するのは困難であるため，RL サーバが送信レート決定時に適切であると判断した送信レートを暫定的に最適な送信レートとする．RL サーバへの報酬の付与は UE がデータ送信後の eNB 負荷状況の変化から，送信レートが帯域利用効率を最大化する適切な送信レー

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

表 1：Q 学習パラメータ

項目	設定値
状態s(状態数:10)	eNB 負荷
行動a	1(送信)/0(蓄積)
報酬 r(送信時)	$+1(L_t \leq L_{ideal})$
	$-1(L_t > L_{ideal})$
報酬 r(蓄積時)	$+1(L_t > L_{ideal})$
	$0(L_t \leq L_{\mathrm{ideal}})$
学習率α	0.1
割引率γ	0.9
ε -greedy	0.1

表 2：評価シナリオ

項目	設定値
UE 数	2 台
シミュレーション時間	600s
遅延耐性時間	600s
eNB 許容負荷	20Kbyte/s
理想負荷	16Kbyte/s(80%)
データサイズ	120Kbyte
データ生成間隔	30s(32Kbps)
データ生成時間	0s～600s

トであるか評価し付与する．再度送信レートを決定する場合，RL サーバが得られた報酬を考慮し送信レートを再決定するため，状態から送信レートを決定し報酬を付与する過程を繰り返すことで，RL サーバが帯域利用効率を最大化する送信レートを獲得する．

4. 基礎評価

4.1 実験条件

提案手法が帯域利用効率を向上させることを確認するため，eNB の負荷変動に応じた送信レート制御を行い，時間的局所性を解消できるかシミュレーションで評価した．評価シナリオは eNB 負荷が時間経過に伴い変化する場面を想定した．**表 1** に提案手法における各パラメータ，**表 2** に評価シナリオを示す．状態sは eNB 負荷とし，状態数を 10 として許容負荷を状態数で分割し割り当てた．行動aはデータの送信と蓄積をそれぞれ 1 と 0 で表現し，報酬rは送信レート制御後の eNB の負荷変動から−1, 0, +1 の報酬を与えることとした．送信レートの決定は，eNB 負荷を制御する目標値(L_{ideal})から eNB 負荷内の遅延負荷データ($L_{realtime}$)を差し引いた帯域を UE 数nで分割し各 UE に割り当てた．式(2)に送信レートRを示す．

$$R = \frac{L_{ideal} - L_{realtime}}{n} \qquad (2)$$

本評価では，学習の収束が確認できた送信レート制御モデルの時間的局所性の解消について評価した．学習の収束を示す指標として，Q 値の平均値を用いることとした．

4.2 実験結果

図 2 にシミュレーション毎の Q 値の平均値を示す．Q 値の平均値がシミュレーション回数の増加に伴い収束していることがわかる．その中でも，Q 値の平均値が最も高いシミュレーション回数 30 回目の eNB 負荷と各 UE の送信レートを**図 3** に示す． eNB 負荷と各 UE の送信量を比較すると，各 UE が eNB 負荷を理想負荷に近づけるように送信レートを制御していることに加え，理想負荷を超過する区間では送信していない．そのため，提案手法が eNB 負荷変動に応じた送信タイミングを学習し，理想負荷に近づけるように制御したと考えられる．一方，eNB

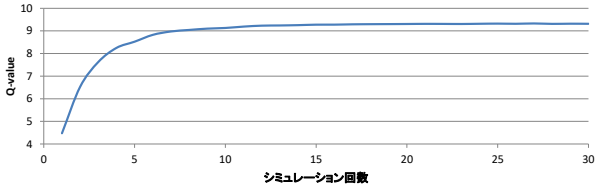

図 2：Q 値の平均値

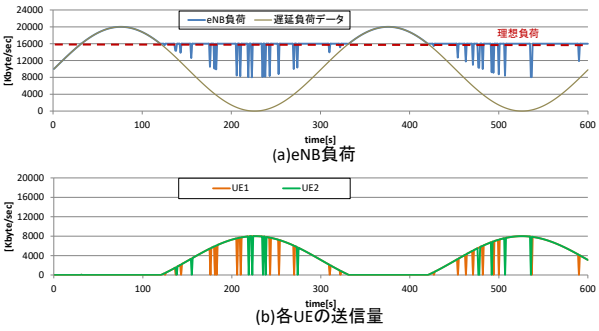

図 3： eNB 負荷と各 UE の送信量

負荷が理想負荷を超過していないが，各 UE の送信レートが安定していない．これは，提案手法が eNB 負荷の帯域に空きがあるにも関わらず送信するべきではないと判断したためである．そのため，Q 学習のパラメータや ε -greedy 法の行動の割合を最適化することで更なる精度向上が見込まれる．以上の結果から，提案手法が時間的局所性を解消し，帯域利用効率が向上したことを確認した．

5. 今後の展開

本稿では，MDOP の時間的オフローディングにおける帯域利用効率の向上を目的とした，強化学習を用いた送信レート自律獲得制御手法を提案した．基礎評価の結果，提案手法を用いることでシミュレーション回数の増加に伴い送信タイミングの精度が向上し，時間的局所性を解消できた．したがって，提案手法が帯域利用効率を向上したといえる．

今後，UE や eNB を増加させ，実環境に基づいた評価を行う予定である．また，送信タイミングだけでなく送信量に強化学習を適用し帯域利用効率の精度向上を目指す．

参考文献

[1] 西岡 哲朗, 他.: モバイルデータオフローディングプロトコル（MDOP）の提案, DICOMO2014 シンポジウム論文集, pp. 613-620(2014).
[2] 3GPP TR 23.705 :User Plane Congestion management (Release-12).
[3] 鈴木理基,他.:LTE 網におけるサービス単位のトラヒック収容技術の検討, DICOMO2014 シンポジウム論文集, pp. 1326-1333(2014).
[4] Y Timmer, et al.: Network Assisted Rate Adaptation for Conversational Video over LTE, Concept and Performance Evaluation, Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing workshop, pp.45-50(2014).
[5] C.J.C.H. Watkins :Learning from Delayed Rewards, Cambridge University PhD thesis(1989).
[6] M Bennis, et al.: A Q-learning Based Approach to Interference Avoidance in Self-Organized Femtocell Networks, Globecom Workshops, pp.706-710 (2010).

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.