

CCN を利用した車車間通信における強化学習によるキャッシュ制御

溝尻 航希, 湯 素華

電気通信大学

1 概要

近年, モビリティを支える技術として車車間通信が注目されている. 同じ道路を通る車両は同じ情報 (以降コンテンツ) を要求することがある. ある車両が要求したコンテンツを後で別の車両が要求する場合, 別々にコンテンツを配信すると効率が悪い. このような繰り返し配信の効率化には, コンテンツ指向ネットワーク (CCN) [1] が知られている. 先行方式[2]では, CCN を車車間通信に適用し, さらにチャネル混雑度に基づいて確率的にキャッシュする手法を提案し, キャッシュの利用効率を上げた. しかしこの手法では, シナリオに適したキャッシュ確率を出力できない.

本検討方式では, キャッシュ利用のさらなる効率化のために, 強化学習を用いて, チャネル混雑度などの情報からキャッシュ確率を制御する手法を検討・評価する.

2 先行方式

CCN とは, 通信ネットワークにおいてコンテンツを要求する際, 配信者の IP アドレス (コンテンツの場所) ではなく, コンテンツの名前を指定するネットワークアーキテクチャである [1].

CCN においてコンテンツの要求者は, 配信者に向けてコンテンツ名を含めた Interest パケットを送信する. Interest パケットを受信した CCN ノードは, 要求されたコンテンツを持っていない場合, 配信者向けに Interest パケットを転送し, 持っている場合, これを Data パケットとして要求者に返送する. この際, Data パケットを中継した CCN ノードはそのコンテンツをキャッシュとして保存する. これにより, 後で同じコンテンツを要求されたときに, キャッシュを利用して応答時間を短縮するだけでなく, 通信量も大幅に削減できる.

[2]では, CCN を車車間通信に適用し, コンテンツの名前付け方式とルーティング方式を考案し, 通信効率を上げることができた. ところが, 各車両のキャッシュバッファは限られているため, 全ての情報をキャッシュすることは不可能である. また, 隣接車両が同じコンテンツをキャッシュすると冗長なキャッシュによりバッファ利用率が低下するため, 適切なキャッシュ制御が必要である. この方式では, 冗長なキャッシュを避けるために Data パケットのキャッシュを確率的に行い, その確率をチャネル混雑度に基づいて制御している. 具体的には, チャネル混雑度が増加した場合はキャッシュ確率を上げ, チャネル混雑度が減少した場合はキャッシュ確率を下げる, という処理を行う.

この制御によってキャッシュ利用効率が向上することが確認された. しかし, この方式の問題点として, シナリオに適したキャッシュ確率を出力できないという点が挙げられる.

3 検討方式

CCN を用いた車車間通信において, キャッシュのさらなる効率化を図るために, 強化学習によるキャッシュ確率の制御方式を検討する.

強化学習 [3]は, 意思決定の主体である「エージェント」とその調査対象である「環境」からなる. エージェントは環境の状態から行動を実行し, 環境が行動によって状態遷移し, 報酬をエージェントに与える. この際エージェントは状態集合から行動集合への写像関数を持ち, これを方策と呼ぶ. エージェントは環境との相互作用を繰り返して, 報酬が最大になるような方策を学習する. 強化学習の利点として, 制御規則の改良を進めることの他に, 複数の要因を取り入れることが可能であるという点が挙げられる.

検討方式では, 強化学習の方式として Q 学習を採用する. Q 学習は, 方策として Q 値と呼ばれる関数 $Q: S \times A \rightarrow \mathbb{R}$ を持つ (S は状態集合, A は行動集合). $Q(s, a)$ は, ある状態 s においてある行動 a を行った際, 現在および将来に得られる報酬の合計になるように学習される. Q 学習とは行動選択の度に Q 値を更新することで, より良い Q 値を

得る学習方法である。

エージェントは、CCN ノードである各車両に搭載されたキャッシュ確率制御システムである。状態集合として、チャンネル混雑度と現在のキャッシュ確率を採用する(表 1)。これらは連続値であるため、性能と学習回数を考慮して離散化する。行動集合はキャッシュ確率の増減である。これにより、エージェントは状態の観測としてチャンネル混雑度とキャッシュ確率を検知し、それに応じたキャッシュ確率の制御方を学習する(図 1)。Q 値の更新間隔を 100 ミリ秒とする。方を最適化するための報酬は、更新の間にキャッシュを利用しなかった場合のコンテンツの送信ホップ数と利用した場合のホップ数の差、つまり節約できたホップ数の合計とする。これが多いほど、チャンネルの利用効率が上がり、要求に対する応答速度が向上するためである。

表 1 強化学習の方針

| | |
|--------|-----------------------|
| エージェント | キャッシュ確率制御システム |
| 状態集合 | チャンネルの混雑度, 現在のキャッシュ確率 |
| 行動集合 | キャッシュ確率の制御 |
| 報酬 | 節約できたホップ数 |
| 更新間隔 | 100ms |



図 1 エージェントと環境の相互作用

4 シミュレーション評価

4.1 シミュレーション条件

表 2 シミュレーション条件

| 項目 | 値 |
|------------|---------------------|
| シミュレータ | Scenargie |
| フィールド | 2000m×2000m |
| 道路の形状 | 200m 間隔の碁盤目状 |
| 車両台数 | 100 台 |
| 車両の初期位置 | ランダムな位置 |
| 車両の速度 | 15-20m/s |
| キャッシュ更新方法 | Least Recently Used |
| シミュレーション時間 | 200 秒 |
| 通信プロトコル | IEEE802.11p |
| コンテンツ要求の間隔 | 1 秒 |
| バッファサイズ | 0 - 300 個 (50 個毎) |

表 2 に、シミュレーション条件を示す。キャッシュバッファサイズを変えてシミュレーションを行った。各コンテンツが要求される頻度は、Zipf の法則 [4] に従う。

4.2 シミュレーション結果

通信経路長の短縮効果を確認するため、コンテンツが要求者に届くまでのホップ数の平均および受信成功率を、バッファサイズを変えながら計測した。結果を図 2 に示す。比較する方式は、先行研究で採用されていたフィードバック制御方式 (ECV+) と、検討方式 (RL/Reinforcement Learning) の 2 つである。

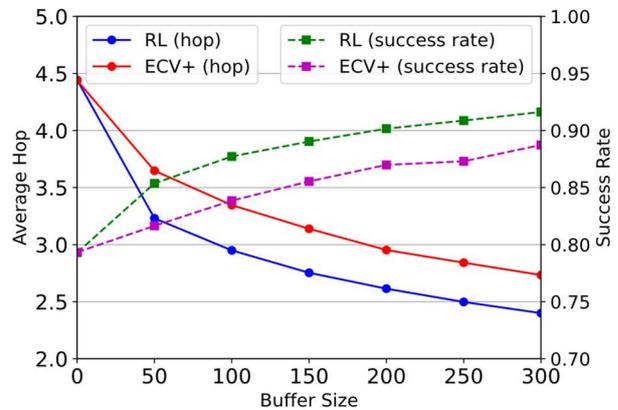


図 2 バッファサイズ毎の結果

バッファサイズの増加に伴って、両方式のホップ数は減少するが、先行方式と比較して検討方式は、ホップ数をさらに約 12% 減少させることができた。また、受信成功率は約 0.05 向上した。適したキャッシュ確率を強化学習によって取得し、キャッシュが効率的に運用されたためだと考えられる。

5 終わりに

本稿では、強化学習によるキャッシュ制御を用いることで、ホップ数の減少が確認できた。今後は、チャンネル混雑度以外の状態を加えた制御手法を検討する予定である。

参考文献

- [1] Van Jacobson, et al., "Networking named content," in *Proc. ACM CoNEXT'09*, pp.1-12, 2009.
- [2] Takanori Nakazawa, et al., "CCN-based inter-vehicle communication for efficient collection of road and traffic information," *Electronics*, vol.9, no.1, 112, 2020.
- [3] 木村 元, 他, "強化学習システムの設計指針," 計測と制御, 38(10), pp.618-623, 1999.
- [4] 佐野 幸恵, "Zipf 則・Heaps 則とその周辺," 情報の科学と技術, 70(2), pp87-89, 2020.