

推定到達時間の局所同期に基づく適応型ルーティングアルゴリズム

里山大河† 平山勝敏† 沖本天太†

†神戸大学大学院海事科学研究科

1. はじめに

Q-routing は, J. A. Boyan らによって提案された, 強化学習を応用したルーティングアルゴリズムである[1]. Q-routing では, 各ノードの学習によってそれぞれが持つルーティングテーブルを更新する. テーブルの学習には, 強化学習の手法の一つである Q 学習を応用しており, 混雑を回避するように学習することで, 迂回路を利用するようになり, 通常ならば負荷が集中しがちなノードの負荷を分散することができる. 本論文で提案する T-routing は, Q-routing のアルゴリズムを強化学習とは別の視点から発展させたものである.

2. Q-routing

本論文では, ネットワーク内のあるノードに発生した荷物やパケット (以下これらをまとめてコンテナと呼ぶことにする) を目的ノードに運ぶ問題を考える. ここで, コンテナはネットワーク内のノードに時刻ごとに確率的に発生する.

ノード v_i が, 目的ノードが v_d であるコンテナを隣接ノード v_j に送ったときの所用時間の推定値を $Q_i(v_j, v_d)$ とし, 強化学習の Q 値に相当する. ノード v_i は, 隣接ノード v_j と目的ノードとなりうるノード v_d の全ての組について, $Q_i(v_j, v_d)$ を持つ.

まず, ノードの待ち行列の先頭にあるコンテナの目的ノード v_d を確認する. コンテナを送るノードを選択する方策には ϵ -greedy 方策を用いる. これは, 確率 ϵ でランダムに行動し, 確率 $1 - \epsilon$ で推定値が最も良い行動をするものである.

ノード v_i が, 隣接ノード v_j にコンテナを送った時にノード v_j のテーブルで最良のもの, つまり

$$Q = \min_{v_k \in \text{neighbor of } v_j} Q_j(v_k, v_d)$$

を問い合わせる. ノード v_i は, 受け取った Q をもとに, 次のように Q 値を更新する.

$$Q_i(v_j, v_d) \leftarrow Q_i(v_j, v_d) + \alpha \{q_i + Q - Q_i(v_j, v_d)\}$$

ここで, q_i はノード v_i における待ち行列の長さ, α は学習率である. つまり, 新しい推定 $q_i + Q$ と既存の推定 $Q_i(v_j, v_d)$ の差を学習率 α によってどの程度反映するかを調整しており, α が 1 に近いほど, 新しい推定を反映して更新する.

Q-routing と強化学習とは, 状態はコンテナの目的ノード, 行動はどの隣接ノードに送るか, 報酬 (費用) は待ち行列の長さそれぞれに対応づけられる.

Q-routing には, その発展アルゴリズムとして, CQ-routing[2], DQ-routing[3], PQ-routing[4]などが存在している. それぞれのアルゴリズムの詳細はここでは省略する.

3. T-routing

T-routing では, テーブルの値を, $T_{min}(v_i, v_d)$ として保持し, これはノード v_i からノード v_d への所用時間の推定値を表す. ノード v_i は, $T_{min}(v_i, v_d)$ と, すべての隣接ノード v_a について $T_{min}(v_a, v_d)$ をもつ. ノード v_i は, 自身の持つ $T_{min}(v_a, v_d)$ を参照して,

$$T = \min_{v_a \in \text{neighbor of } v_i} T_{min}(v_a, v_d)$$

を求め, その T を与える隣接ノードにコンテナを送る. その後, Q-routing と同様に, 以下の更新式でテーブルの値を更新する.

$$T_{min}(v_i, v_d) \leftarrow T_{min}(v_i, v_d) + \alpha \{q_i + T - T_{min}(v_i, v_d)\}$$

更新した後, その更新した $T_{min}(v_i, v_d)$ をすべての隣接ノードに通知する.

Q-routing では, 問い合わせたときのみ Q 値を更新していたため, Q 値がネットワークの最新の状態を反映していない可能性がある. 一方, T-routing では, 各ノードは, 自分自身の隣接ノードがテーブルの値を更新する度にその値に同期するため, コンテナを送るノードを決定する際に常に最新の値を利用できる. また, その値を利用して更新した値もまたネットワークの最新の状態を表している可能性が高い.

T-routing は, Q-routing と異なり, 状態の観測, 行動, 報酬の獲得の順番ではないため, 強化学習とは異なる. T-routing では同じノードを選択し続けることによって混雑が発生すると, 相対的に他のノードの推定値の方が小さくなる. これによって選択の変更が促され, これを混雑が発生する度に繰り返しているため, 負荷分散を行える.

4. 実験

図 1 に示すようなネットワークを対象に実験を行う. このネットワークでは, ノード間の移動において, ほとんどの場合中央付近の \times 印を入れた

Adaptive routing algorithm based on local synchronization of estimated arrival time

Taiga Satoyama, Katsutoshi Hirayama, Tenda Okimoto
Graduate School of Maritime Sciences, Kobe University

エッジを利用するのが最適であるため、周辺のノードがボトルネックとなり、混雑の原因となる。そのため、上の迂回路を利用することで負荷を分散することが求められる。

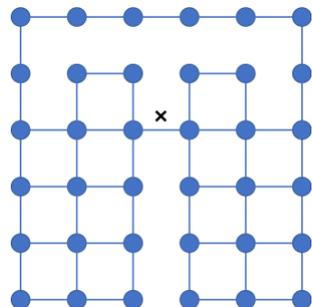


図 1. 実験対象としたネットワーク

パケット発生確率が 0.05 の時を低負荷時, 0.1 の時を高負荷時として, 最短経路に固定する場合, Q-routing, CQ-routing, DQ-routing, PQ-routing と比較を行う. 各時刻における待ち行列の長さの最大値, つまり最も混雑しているノードの待ち行列の長さの推移を調べた.

図 2 は, 低負荷のときの推移である. Q-routing は初期の学習期間では混雑が発生しているが, 学習が十分進むと, 最短経路の場合と同程度の性能

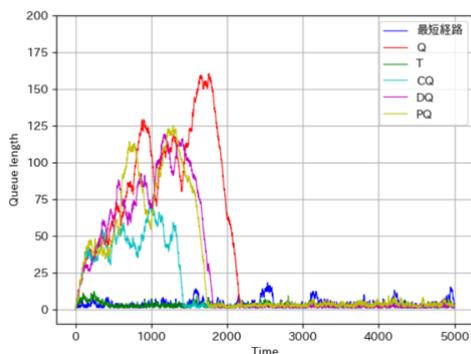


図 2. 低負荷時の結果

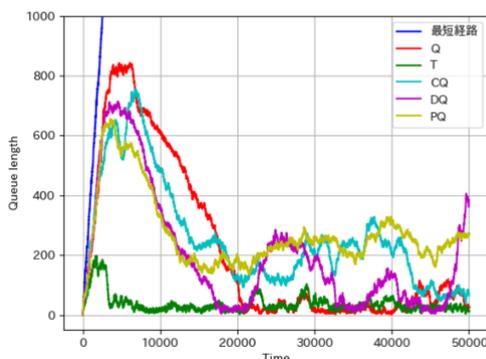


図 3. 高負荷時の結果

を發揮する. これは CQ-routing, DQ-routing, PQ-routing でも同様であり, Q-routing と比較すると初期の学習期間がより早く収束している. T-routing は, 初期の学習期間における混雑がほぼ発生しておらず, 最短経路を採用した場合と遜色ないことがわかる.

図 3 は, 高負荷のときの推移である. 高負荷のときは初期の学習期間で混雑が見られるが他の発展アルゴリズムと比較しても小さく, 収束も早い. 学習期間終了後においても T-routing は大きな混雑が見られない.

5. まとめ

本論文では, Q-routing の発展アルゴリズムとして, T-routing を提案した. T-routing では, Q-routing を参考に隣接ノード間での情報の同期を行うことで情報を効率的に伝播し, 学習を効率化した. 実験の結果, 他の発展アルゴリズムと比較しても初期の学習期間が早く収束し, 高負荷であっても十分な性能をもつことがわかった.

今後の課題として, 実際の情報ネットワークや物流ネットワークを対象としたケーススタディを行うことなどが挙げられる.

参考文献

- [1] J.A. Boyan & M.L. Littman (1993). A distributed reinforcement learning scheme for network routing. Proceedings of the First International Workshop on Applications of Neural Networks to Telecommunications.
- [2] Shailesh Kumar & Risto Miikkulainen (1998). Confidence-Based Q-Routing: An On-Line Adaptive Network Routing Algorithm. Proceedings of the Artificial Neural Networks in Engineering Conference.
- [3] Shailesh Kumar & Risto Miikkulainen (1998). Dual Reinforcement Q-Routing: An On-Line Adaptive Routing Algorithm. Proceedings of the Artificial Neural Networks in Engineering Conference.
- [4] Samuel P.M. Choi & Dit-Yan Yeung (1996). Predictive Q-Routing: A Memory-based Reinforcement Learning Approach to Adaptive Traffic Control. Advances in Neural Information Processing Systems.