5J-04

Q学習による相互結合網制御の初期検討

深澤 勇人 † 横田 隆史 †† 大津 金光 ††
† 宇都宮大学工学部情報工学科 †† 宇都宮大学大学院工学研究科情報システム科学専攻

1 はじめに

大規模な並列計算機システムは，多くの演算ノードで構成されるため，ノード間における通信性能は系全体の計算器性能に大きな影響を及ぼす．並列計算機の性能を高めるために，通信部分を担う相互結合網の検討は必須である．しかし相互結合網には，過大な通信負荷により輻輳状態が生じる結果，転送性能が著しく低下する，という問題がある．

この問題に対処するための輻輳制御手法の 1 つに，パケットの投入を抑制する手法がある [1]. これは，パケットの投入タイミングを適切に管理することで，輻輳が生じずにいられる領域内でパケットの流量が最大になるよう制御する，という発想に基づく．これによりパケット投入の適切な抑制は，輻輳発生による性能悪化に効果的に作用することが明らかとなった．そこで本論文では強化学習の 1 手法である Q 学習 [2] を用いて，パケット投入の適切な抑制を学習させることを目指す．その初期検討として，選択した 1 ノードにのみ学習を適用した結果を示す．

2 Q学習の適用の検討

相互結合網における輻輳の発生は，パケット間で干渉し起こった転送の妨げが，別の転送の妨げに繋ってしまうため起こる．つまり輻輳を避けるようパケット投入のタイミングを学習するには，系全体のトラフィックを常に監視し，それに依った判断を下す必要がある．このような系全体の状態を見て，取る行動を適切に学習するのに適しているのが強化学習である．

Q 学習は試行錯誤を繰り返すことで，最大の価値を得られるよう学習していく強化学習の一つの手法であり，1 つのデータに対して正解不正解が分からずとも評価を得て学習を進められるという特徴がある．これは囲碁や将棋といった盤面ゲームを学習する AI の一部に使われている．とある巡目のとある指し手だけを見て評価する事は人間でも難しい．しかし Q 学習では最終的な勝敗から遡ることでそれぞれの指し手に評価を与えることができる．

Q 学習による相互結合網制御の具体的な方法を説明する．プロセッサがパケットを生成し送信するタイミングに，Q 学習による判定を挟む．この判定部では，生成したパケットを，送信する/待つの 2 つから選択し行動する．学習を繰り替えすことで，状況に応じて送信するか待つかを正しく判断できるようになることを期待する．その初期検討として，選択 1 ノードにのみこの学習を適用する．

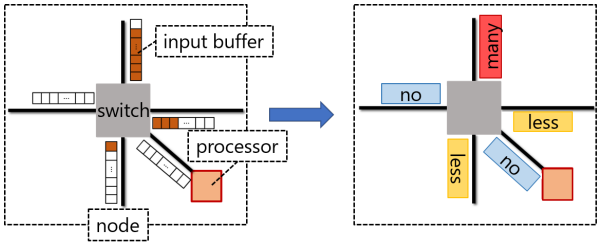

図 1: バッファの埋まり具合を縮約するイメージ

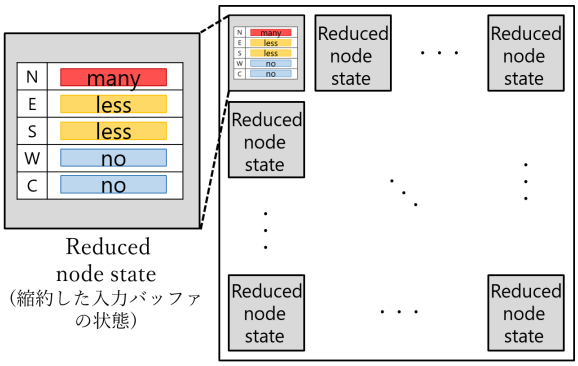

図 2: ネットワーク状態 S_n

Q 学習を本研究用に設計するために，“状態”，“行動” “報酬” の 3 要素を定める．まず “状態”について考える．Q 学習において “状態” は現在がどんな状況であるかを示したもので，外部から得られる情報の全てであると言える．本学習では相互結合網全体のトラフィック状況を見なければいけないので，ここでは各サイクルタイムにおける，全ノードの入力バッファの埋まり具合を “状態” と定めたい．しかしこれでは想定する状態数が多いという問題がある．

そこで入力バッファの埋まり具合を簡素化することで状態数の低減を図る．各入力バッファの埋まり具合を，フリットが 1 つもない (no)，半分以下 (less)，半分以上 (many) の 3 つに分類する．そうして表した全体の様子を，新たに “状態” とすることで状態数を減らす．各入力バッファの埋まり具合を, 前述した 3 つに分類する縮約のイメージを図 1 に示す．この新たに示した “状態” を以降，ネットワーク状態と呼称する．ネットワーク状態のモデルを図 2 に示す．次に “行動” は送信する/待つの 2 つとする．“報酬” は指定回数の一斉同期通信が終了するまでにかかった時間に比例して与える．当然かかった時間が短い程与える報酬は大

Preliminary evaluation of interconnection network control by Q-learning
†Yuto Fukasawa,††Takashi Yokota,††Kanemitsu Ootsu,
Department of Information Science, Faculty of Engineering, Utsunomiya University (†)
Department of Information Systems Science, Graduate School of Engineering, Utsunomiya University (††)

きくなるようにし，比例させる関数は単純に一次関数とした．

次に学習の流れを簡単に説明する．シミュレーションが始まると，毎サイクル毎に表れたネットワーク状態を全て保存していく．ここで保存した各ネットワーク状態に対して，送信する/待つに対応するQ値が初期値ランダムで与えられる．選択したノードのプロセッサがパケットを生成し送信しようとしようとすると，その時のネットワーク状態から対応するQ値を呼び出しそれに応じて送信する/待つの行動を決める．これをQ値判定と呼称する．

判定を行ったネットワーク状態における行動は，行動列として記録しておく．指定回数の通信が終わった際，この行動列を使ってQ値の更新を行う．報酬が与えられるとそれに応じて行動列を遡りQ値に与えていく．つまり取った行動のQ値のみが更新される仕組みである．指定回数の通信が終わりQ値の更新が行われることを1回の学習とする．学習の回数を重ね十分な回数, Q値の更新が行われたすると，それぞれのネットワーク状態において適当な行動が取れるようになると考えられる．

3 学習結果

相互結合網シミュレータに前章で示した機能を実装した．トポロジは8x8のトーラス網を想定する．パケット長は8フリットで，入力バッファの容量は16フリット分，仮想チャネル数は3とした．ルーティングアルゴリズムは固定型である，次元順ルーティングを用いる．通信パターンはビット反転，ビット逆順，ビット回転，転置から選び，選択ノードはランダムで選択し様々な組み合わせで試した．各ノードが5つのパケットを送出する一斉同期通信にかかった時間を計測する．学習は2000回行った．

まずパケット毎にQ値判定を行い実験をした．その結果全ての通信パターンにおいてサイクルタイムの変化はほとんど表れなかった．5つのパケットを送るのでQ値判定は5回起こる．つまりこれは64ノードある中の1ノードで, 5回パケットの送信をずらした程度では全体のサイクルタイムにほとんど影響がでないということを示している．

次にフリット毎にQ値判定を行うように変更し実験を行った．その結果サイクルタイムが大きく変化した．この研究におけるQ学習の特徴として，同じネットワーク状態が連続して表れることが判明した．これは状態を縮約していることと，行動に “待つ” があるためである．ここに起きる問題としてQ値更新の回数に偏りが生まれることがわかった．そのため, Q値に差が生まれすぎてしまい，報酬による更新が正しく機能していないことが考えられる．そのため同じネットワーク状態が連続した場合, Q値更新は一回のみにするという制限をかけるよう修正を行った．

制限をかけた結果を図3に示す．通信パターンはビット反転で，ノード (3,3) を選択したものである．図3

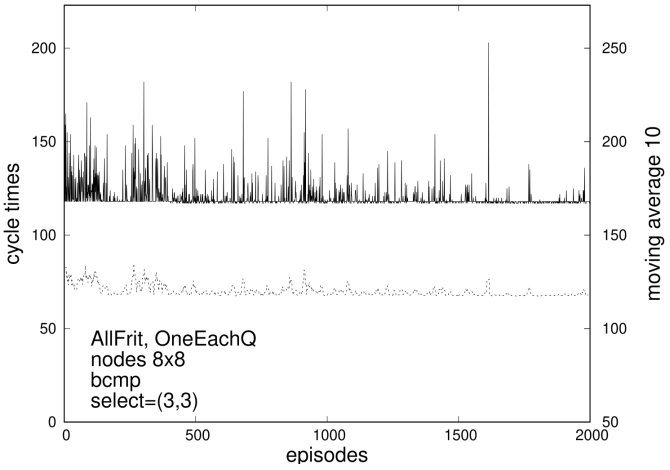

図 3: 学習回数に対するサイクルタイムの推移

は横軸に学習回数をとり，縦軸はそれに対するサイクルタイムを示している．実線はサイクルタイムを示したもので値は縦軸右を，点線は幅10で移動平均をとったものをずらして示しており，値は縦軸右を見る．傾向として学習回数が増える程サイクルタイムぶれ幅が小さくなっている．500[cycletime] あたりで最低値が更新され，以降その値に引っ張られるように値が出現している．頻繁に表れるサイクルタイムが1縮まった．これはサイクルタイムが短くなるよう学習している証拠だといえるだろう．

他の結果に関して，他ノードを選択した場合や通信パターンを変えた場合の結果は，ある程度収束するよう学習したパターンと，そうならずひたすらぶれるパターンに分かれた．ここに生まれた差は，通信パターンと選択したノード毎に存在する，経路の重なりによるものだと考えている．

4 おわりに

本稿では, 相互結合網における輻輳の発生を抑えるためのパケット投入の適切な抑制について, Q学習を用いる手法を提案し初期検討をした．選択した1つのプロセッサのパケット送信処理を学習に沿わせ，いくつかの通信パターンと複数の選択ノードでシミュレートした結果，いくつかの組み合わせで1サイクル早く通信が終わった．今後の進展としては，複数ノードにおける本手法の実装が第一である．

謝辞

本研究は，一部日本学術振興会科学研究費補助金 (基盤研究 (C)16K00068，同 (C)17K00265) の援助による．

参考文献

[1] 横田 隆史, 大津 金光, 古川 文人, 馬場 敬信: “エントロピー・スロットリング:相互結合網のパケット移動度に着目した輻輳制御手法”, 情報処理学会論文誌:コンピューティングシステム，Vol.47，No.SIG 12(ACS 15)，pp.1-11，2006年9月.

[2] C.J.C.H. Watkins: “Learning from Delayed Rewards,” PhD thesis, King’s College, University of Cambridge, 1989.