大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

トップエスイー ソフトウェア開発実践演習

Q学習による有向グラフ上の特定ノードの利用効率最適化手法の検討

(株)日立ハイテクノロジーズ　森月 政博　masahiro.morizuki.wz@hitachi-high-tech.com

開発における問題点

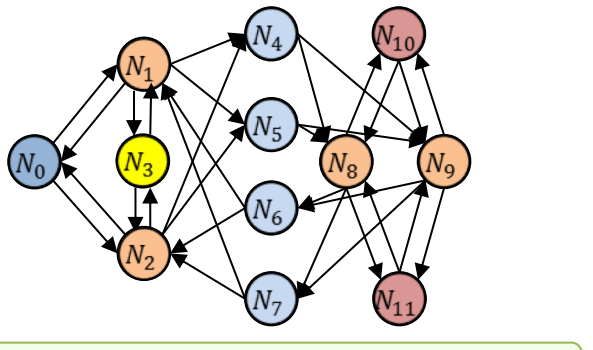
有向グラフ上を移動するメッセージが，ある特定のノード(処理ノード)上で処理されるモデルにおいて，処理ノードの利用効率を最大化するメッセージ移動経路を割り出す必要がある．しかしグラフが取り得る状態の数が膨大であるため，通常の最適化手法的アプローチでは状態爆発が起こり，有限時間内の解の導出が困難．

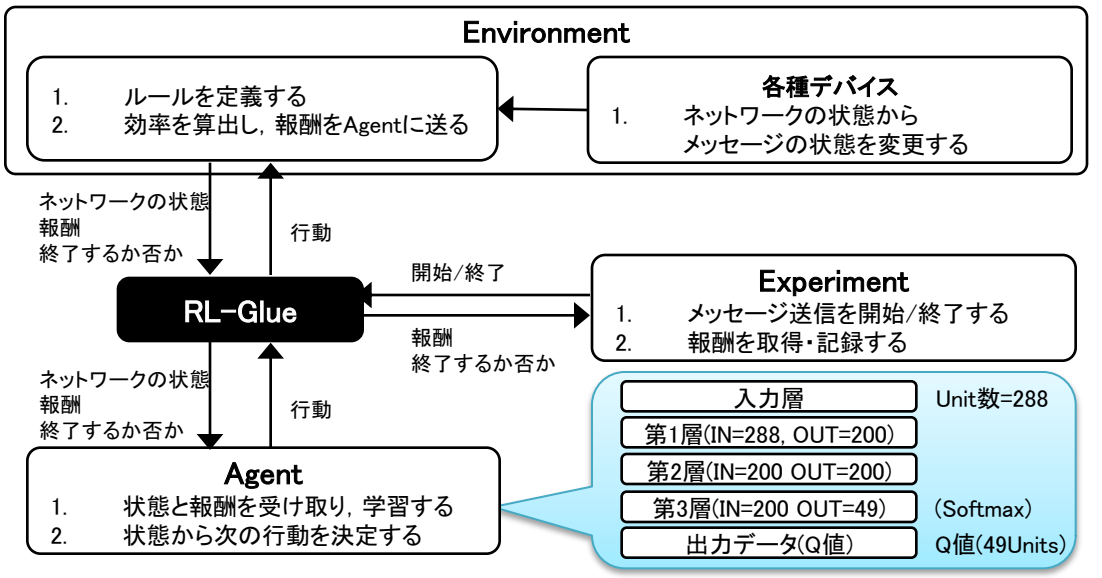
手法・ツールの適用による解決

学習した状態の近似から解を予想するDQN(Deep Q-Network)を用いて、決められたルールに従って広大な状態空間内での最適な状態遷移を導くシステムを構築する。処理ノードの利用効率を最大化するための指標を選定し、システムの構築・評価を行った。

モデルと構成

対象とする有向グラフ(ネットワーク)





実験・評価

報酬の獲得ルール

処理ノード利用効率(%)	報酬
40≦f(P, M)	+1.0
30≦f(P, M)＜40	-0.2
f(P, M)＜30	-0.6
ルール違反等のエラー	-1.0

$$f(P,M)=\frac{\sum_{i=1}^{|P|}\sum_{j=1}^{|M(P_i)|}Pt(M(P_i)_j)}{|P|\times T}\times 100$$

P: 使用した処理ノードの集合
M: 送信されたメッセージの集合
T: 算出基準時間
$M(P_i)$: ノードP_iで処理したメッセージの集合
$Pt(M(P_i)_j)$: メッセージ$M(P_i)_j$の処理時間

Experiment	F(P, M)	学習完了ステップ数
初回	24%	4500
ルール追加①(Environment変更)	24%	3200
ルール追加②(Environment変更)	33%	5500
行動追加(Agent変更)	42%	37600

- **評価**
  適切なルールを環境に設定することにより、処理ノード利用効率を改善するよう学習させることができた。
- **考察**
  Agentの変更は効果が大きいが、学習時間に与える影響も大きい。
  学習の効率は報酬の与え方に大きく左右されそう。

今後の取り組み

- **報酬の与え方の再検討**
  今回は学習をスタートさせてから5メッセージを送信完了した時点での処理ノード利用効率を基に報酬を与えることとしたが、報酬を得るまでの時間が長く、初期状態に報酬が伝播するまでの学習時間が長くなった。各ステップ毎に報酬を与えることを検討する。

- **目的指標の見直し**
  今回は単位時間あたりの処理ノードの利用効率を最大化する機械学習プログラムとしたが、これはネットワーク全体のスループットを向上させることに必ずしも寄与しない。今後はスループット向上を目標とし、待ち行列中の在庫滞留を無くすよう個々のノードの利用率を平準化する(利用率の差を最小化する)等を指標とした機械学習も試行したい。