

Actor-Critic を用いた知的ネットワークシステム
中村 康昭

1 はじめに

我々は、複数の知的人工物をネットワークに接合したシステムの総称を「知的ネットワークシステム」と定義し、検討を行っている。知的人工物単体では有限のセンサ部およびアクト部に使用に限られるが、知的ネットワークシステムにおいては、他の知的人工物のセンサ部やアクト部がネットワークを経由し使用可能となるため、システムの可能性はセンサ部とアクト部の組み合わせの数となり拡大する。

2 知的ネットワークシステム

我々は知的ネットワークシステムにおいて、究極には個々の知的人工物のプラグアンドプレイ、設定されていない要求への対応を可能にすることを目標としている。そのためには通信プロトコルの設定、要求の理解とそれに対応したシステムの構築、対故障性など解決すべき問題はいくつも存在するが、その中でも大きな問題の一つが、判断部の設計である。

例えば「要求者が満足するような部屋」という問題に対しては、満足する明るさに関するパラメータを決定する必要がある。あらかじめその値を設定しておくことも可能であるが、要求者の変更、時間、場所の依存性、その他の問題への対応可能性を考慮すると、現実的ではない。そのため本研究では、そのパラメータに学習を利用している。これにより、Fig. 1 に示すように、知的人工物が行動した結果、環境が変化し、その変化を別のセンサで取り込み、判断し、知的人工物の判断部のパラメータを変更するという二層の構造を持つこととなる。

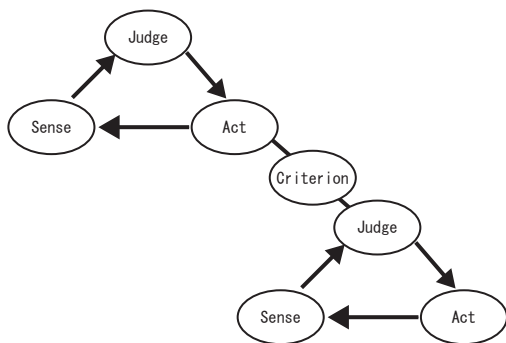


Fig. 1 Elements of Intelligent Artifact using learning

3 強化学習

具体的に用いる学習手法の検討を行う。コンピュータに学習をさせることを目的とするものは機械学習であり、その中で、試行錯誤を基に環境に適應する学習制御の枠組みが強化学習である。強化学習では教師有り学習と異なり、入力に対する理想的な出力を提示する教師が存在しない。その代わりに報酬を環境から得ることにより学習を進める。現在、強化学習の中で TD 誤差学習と呼ばれる方法が主に用いられており、本研究ではこの中の Q-Learning と Actor-Critic について検討を行った。

4 学習手法による比較

本研究では Q-Learning と Actor-Critic の比較実験を行った。実験環境としては、知的ネットワークシステムの一つとして、知的家電ネットワークを取り上げた。ここでネットワークに接続する家電は照明とし、目標を「人のいる地点を快適な明るさにせよ」とし、今回はこの快適な明るさを $100 \pm 5[lx]$ とした。Q-Learning では、 $\pm 20[cd]$ という行動から選択させ、Actor-Critic では正規分布を用いて連続値を扱う。2つの学習手法によって動作基準を獲得させた結果を Fig. 2 に示す。

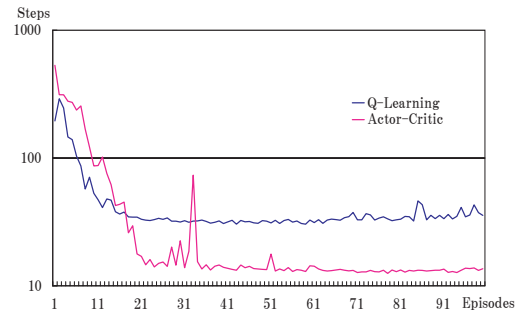


Fig. 2 Comparison of Q-Learning and Actor-Critic

Fig. 2 より、Actor-Critic の方が学習が進むとより少ないステップ数で目標状態に到達することが分かる。

5 まとめ

本研究では、知的ネットワークシステムに用いる学習手法の検討を行った。

シミュレーション結果より、連続的な行動出力が要求される時、Q-Learning よりも Actor-Critic が有効であることが確認できた。