Actor-Critic を用いた知的ネットワークシステムの提案
A proposal of Intelligent Network System using Actor-Critic

同志社大学工学部　　廣安 知之，三木 光範，○ 中村 康昭
Tomoyuki HIROYASU, Mitsunori MIKI, and Yasuaki NAKAMURA
Department of Knowledge and Computer Science, Doshisha University

1　はじめに

我々は一連の研究で複数の知的人工物をネットワークに接続したシステム,「知的ネットワークシステム」の検討を行っている [1, 2, 3]．これまでの研究を通じて，知的ネットワークシステムの知的レベルを向上させるためには，システムの学習能力が不可欠である事がわかった [2]．一般に，人工物の状態遷移においては離散的な場合よりも連続的な場合が多い．そこで，本研究では連続的な場合における学習方法として，Q-Learning と Actor-Critic を取り上げ，両手法の比較実験を通じて Actor-Critic の優位性について検討を行った．

2　知的人工物

人工物とは，人によって作られた“ もの ”の総称である．近年の人工物の中には，人および環境の負担を軽減するため自身の制御を行う人工物が存在する．例えば，マイコン制御の炊飯器は温度認識などから，熱加減の調整を行う事により利用者の負担を軽減している．

我々はそのような人工物を知的人工物と定義し，その知的性質に関する研究を行ってきた．センス部，ジャッジ部およびアクト部を有し，環境に合わせて選択肢の中から最適な行動を判断，動作する．この一連の動作を図 1 に示す．

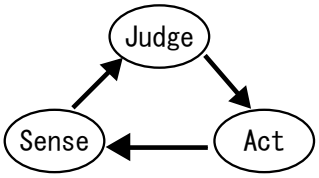

図 1: Behavioral dissolution of intelligent characteristics in artifacts.

3　知的ネットワークシステム

近年，ネットワーク利用の普及に伴い，あらゆる人工物がネットワークに接合されるようになってきた．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」と呼んでいる．知的人工物単体では有限のセンス部およびアクト部使用が限られるが，知的ネットワークシステムにおいては，他の知的人工部のセンス部やアクト部がネットワーク越しに使用可能となるため，システムの可能性はセンス部とアクト部の組み合わせの数となり拡大する．

我々は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，満足する明るさがいくらかという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．こうすることで，図 2 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有することとなる．

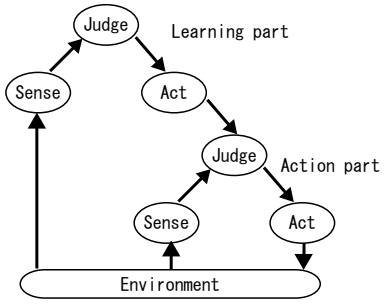

図 2: Two levels of intelligent elements using learning

4　強化学習

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応する学習制御の枠組みが強化学習である [4]．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．

4.1 TD 誤差学習

TD 誤差学習では，現在の状態がどの程度よい状態であるかを見積もり，その状態に対する評価値を与える．そして実際の行動により，評価値が正しいかを確認する．評価値の更新は式（1）に従う．

$$V(s_t) \leftarrow V(s_t) + \alpha \times TD\text{-}error \qquad (1)$$
$$TD\text{-}error = r + \gamma V(s_{t+1}) - V(s_t)$$

式（1）において，$V(s_t)$ は時刻 t における状態 s の評価値，r は報酬，α は学習率，γ は割引率を示す．

4.2 Q-Learning

TD 誤差学習の一つである Q-Learning では，状態と行動の組に対する評価を見積もる．その見積もった評価値から，その状態における適した行動の判断を行う．ある状態で行動し，見積もりよりも良い評価値を持つ状態に遷移をした時には選択された行動の評価値が上昇し，逆に遷移した状態が見積もりよりも悪いときには行動の評価値は減少する．評価値の更新は式 (2) に従う．

$$V(s_t, a_t) \leftarrow V(s_t, a_t) + \alpha \times TD\text{-}error \qquad (2)$$
$$TD\text{-}error = r + \gamma \max_a V(s_{t+1}, a) - V(s, a_t)$$

つまり TD 誤差学習の状態評価値の更新において，遷移後の状態における行動の評価値のうち，最大の評価値を遷移後の状態評価値とし，状態と行動のセットに対して行動の評価値を更新する．

4.3 Actor-Critic

Q-Learning に対し，Actor-Critic では状態評価部と行動選択部が独立した形で存在する．行動選択部には確率が設定され，連続行動空間では正規分布に基づく方法などが用いられる．本研究では行動選択部にこの正規分布を用いた．各状態ごとに状態の評価値 ($V(s)$)，行動の確率を定める正規分布の中心値 (μ) および標準偏差 (σ) を持つ．行動の選択では各状態の μ と σ に基づいて行動を決定する．行動の結果，TD 誤差が正，すなわちよい状態に遷移したときには，正規分布の中心値を，その行動を選ぶ確率が高くなる方向へ移動させる．また，正規分布の幅に関しては，よりよい状態への遷移が標準偏差の内側で行われた場合には，標準偏差の値を小さくし，よりよい状態へ遷移する行動を選択する確率を高くする．

5 学習手法による比較

本研究では前章において説明した 2 つの学習手法，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 100[lx] の明るさにせよ」とした．Q-Learning では，±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図 3 に示す．

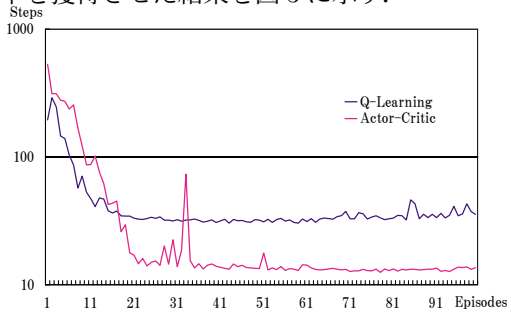

図 3: Comparison between Q-Learning and Actor-Ciritic

図 3 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となる状態に到達していることが分かる．これは，Actor-Critic が実際の行動からその判断基準を生成しているのに対して，Q-Learning はあらかじめ行動を設定しているためである．

以上より，今回のように連続的な行動出力が望まれるような場合は Actor-Critic が有効であると言える．

6 おわりに

本研究では強化学習を用いた知的ネットワークシステムの提案を行った．また，学習方法の実装として，Q-Learning と Actor-Critic を取り上げ，両手法の比較実験を行った．結果より用いる学習手法としては連続的な行動を必要とするときには Actor-Critic が有効である事を確認した．

参考文献

[1] 廣安知之, 三木光範, 冨田浩司. 知的人工物のネットワーク化によるシステムの知的化（知的照明システムの構築）. 日本機械学会第 9 回設計工学・システム部門講演会講演論文集, pp. 518–521, 1999.

[2] 冨田浩司. 知的ネットワークシステムへの強化学習の適用（q-learning による知的照明システムの構築）. 計測自動制御学会 第 13 回自律分散システム・シンポジウム資料, pp. 27–32, 2001.

[3] 中島史裕, 廣安知之, 三木光範. Its における知的ネットワークシステムの構築（- 知的信号機システムの提案-）. 計測自動制御学会 第 13 回自律分散システム・シンポジウム資料, pp. .33–38, 2001.

[4] 木村元・宮崎和光・小林重信. 強化学習システムの設計指針. 計測と制御, Vol. 38, No. 10, 1999.

出典：
第 46 回 システム制御情報学会研究発表講演会論文集
pp. 587-588
(2002 年 5 月)

問い合わせ先：
同志社大学工学部/ 同志社大学大学院工学研究科
知的システムデザイン研究室
(http://mikilab.doshisha.ac.jp)