Actor - Critic アルゴリズムによる 感覚尺度を学習する照明制御システム†

廣安 知之 *1・中村 彰之 *2・三木 光範 *2・吉見 真聡 *2・横内 久猛 *1

本研究では、照明制御システムにおいて「とても明るく」や「やや暗く」のようにユーザが感覚的に明るさを指示し、その尺度を強化学習手法の一つであるActor-Criticアルゴリズムを利用して学習するシステムを構築する。このシステムではユーザの要求とユーザ周辺の明るさの2種類の状態に応じた行動を取るように学習する必要がある。本研究では効率良く学習を行うために、2種類の状態それぞれに対応した2種類のActorを備えたActor-Criticアルゴリズムを構築した。本稿ではそのアルゴリズムの概要とその有効性を示す。

キーワード：照明，感覚尺度，Actor - Critic アルゴリズム

1. はじめに

従来の照明制御システムは複数の照明を一括して制御するシステムであった。一方、現在は複数の照明を個別に制御でき、多様な光環境を作ることが可能となっている。更に、ソフトウェアの進歩により照明の知的な動作も実現されている[1, 2]。しかし、ユーザにとっては操作できる対象が増加かつ高度化することで操作負担が増加している。例えば、複数の照明の明るさを個別に設定する操作には多くの手間が必要となり、照度(物体を照らす明るさの指標：lx)などの一般的でない尺度に基づいて明るさを指定する操作には特殊な知識が必要となる[3, 4]。

この操作の複雑化に対して、「とても明るく」や「やや暗く」といった感覚的な指示により、任意に明るさを操作できるシステムが有用であると考えられる[5, 6]。なぜならこの操作であれば手間は少なく、特殊な知識も必要としないからである。ただし、この「とても」や「やや」という感覚尺度はユーザによって異なる[7]。そこで、システムがユーザごとに感覚尺度を学習する必要がある。本研究ではその学習に強化学習の一手法であるActor-Criticアルゴリズムを用いる。なぜなら、この手法であれば学習に予備実験を必要としないためにユーザの負担を省け、且つ照明の明るさをどれ程度変

化させるかという連続値を扱えるためである[8]。

ここで考慮すべきは、その時の照度に応じて照度の変化量に対する人の感覚量が変わることである[9, 10]。そのため、システムは「とても明るく」や「やや暗く」などの要求をユーザから与えられた時、その要求と要求時のユーザ周辺の照度という2種類の状態に応じて照度を変化させなければならない。2種類の状態を学習する時、従来のActor-Criticアルゴリズムは学習器を1種類しか備えていないため、それらの状態の全組み合わせに対して学習を行うことになる[11, 12]。これでは状態数が増える程、学習に時間が掛かってしまう。本システムにおいては学習時間が増えることはユーザの負担が増えることに等しいため、効率的な学習アルゴリズムが必要となる。そこで本稿では、2種類の学習器を持つActor-Criticアルゴリズム(Two-Actor-Criticアルゴリズム)を備えた感覚的操作による照明制御システムを提案する。

2. 感覚的操作による照明制御システム

1章で述べたように、照明制御システムにおけるユーザの操作負担が増加している問題に対して、「とても明るく」や「やや暗く」といった感覚的な指示によってユーザが明るさを操作できるシステムの構築を検討する。本章ではその感覚的操作による照明制御システムの概要と要件を述べる。

2.1 概要

本システムの概要図を図1に示す。

図1に示す様なコンピュータのGraphical User Interface (GUI)によりユーザはどれだけ明るさを変化さ

† The Sensuous Lighting Control System using Actor-Critic Algorithm

Tomoyuki HIROYASU, Akiyuki NAKAMURA, Mitsunori MIKI, Masato YOSHIMI and Hisatake YOKOUCHI

*1 同志社大学 生命医科学部
Doshisha University, Faculty of Life and Medical Sciences

*2 同志社大学 理工学部
Doshisha University, Department of Science and Engineering

図1 システムの概要図

せたいかを指示する。現在はコンピュータに対して現在の照度からの変化の指示をコンピュータに対して行うが、将来は部屋に存在する家庭用ロボットに対して行うなどの可能性が考えられる。この時、ユーザはシステムが指定する選択肢、「とても明るく」「明るく」「やや明るく」「ほんの少し明るく」「ほんの少し暗く」「やや暗く」「暗く」「とても暗く」などの中から明るさに対する要求を選択する。システムはユーザの要求に基づいて室内の明るさを変更する。要求を入力してはシステムが明るさを変化させるという手順を、ユーザが明るさに満足するまで反復する。ユーザはこのシステムが制御する照明の設置された部屋に入室する度にこのやり取りを繰り返す。初期段階ではユーザの希望となる照度を満たすまでに何度も指示を行わなければならないが、システムはユーザの感覚尺度を学習するので、入室を繰り返すごとにシステムに指示する回数が減少する。また、実際の利用を考えるとユーザが複数人存在することが考えられるが、本研究ではその基礎的な検討としてユーザが1人のみの状況を学習対象とする。

2.2 要件

本システムの要件を以下に示す。

● 照度制御

従来、ユーザはスイッチやつまみにより照明の光度(照明が放つ明るさの指標: cd)を調節していた。しかし、ユーザにとっては照明自体の光度を変化させるのではなく、自らの手元や周囲の照度を変化させることが望ましい。そのため、本システムは部屋の明るさを光度ではなく照度に基づいて操作する。ただし、システムが直接制御できる

ものは照明の光度のみであるから、照度センサを用いて光度と照度の関係を推測することで照度制御を実現する。具体的な照度制御アルゴリズムに関しては、既存研究で用いられているアルゴリズムを利用する[13, 14]。

● 感覚尺度の学習

ユーザの指示である「とても」や「やや」といった感覚的な尺度はユーザによって異なるため、ユーザごとに感覚尺度を学習する必要がある。また、その時の照度に応じて照度の変化量に対する人の感覚量が変化する。そのため、システムはユーザからの要求と要求時のユーザ周辺照度という2種類の状態に対応した学習を行う。本システムでは学習の進行度合はユーザの要求入力回数に比例することから、ユーザの操作負担を減らすために効率の良い学習アルゴリズムが必要となる。

上記の2つの要件の内、本稿では感覚尺度の学習について3章で説明する。

3. 感覚尺度の学習

本研究では感覚尺度を学習するアルゴリズムとして、動的に変化する人の感覚にも動的な学習によって対応できる強化学習を用いる。更に、本システムに求められる出力は照度の変化量という連続値であるため、強化学習の中でも連続値の出力に適しているActor-Criticアルゴリズムに着目する。本システムにおいてはユーザの負担軽減のために効率的な学習アルゴリズムが求められるため、Actor-Criticアルゴリズムを改変したTwo-Actor-Criticアルゴリズムについて検討する。

本章では、強化学習及びActor-Criticアルゴリズム、Two-Actor-Criticアルゴリズムについて説明する。

3.1 強化学習

強化学習とは、システムが試行錯誤を繰り返すことで環境の状態に応じた適切な行動パターン(方策)を学習する手法である[15, 16]。強化学習には環境とエージェントという概念が存在する。環境とは学習における対象問題であり複数の状態を持つ。エージェントは学習を行う存在であり、環境の状態に応じた行動を自らの方策により決定する。エージェントには学習するための指標が必要であるから、環境はエージェントにその行動の評価として報酬を与える。そして、エージェントは将来に渡って得られる報酬の最終的な累積が最大化するように方策を更新していく。本問題ではユーザは環境に含まれ、報酬はユーザの要求の変化に

従って決定される。例えば、ユーザから「とても明るく」という要求を受けてエージェントが照度を変化させ、その次の要求が「ほんの少し明るく」に変わったとする。その時、エージェントの行動によってユーザ周辺の照度が目標照度に近付いたことを表すため、環境はその行動に対して正の報酬を与える。

強化学習では環境のダイナミクスをマルコフ決定過程(MDP)によってモデル化する。MDPとは将来における事象の起こる確率は過去のいかなる状態にも依存せず、現在の状態にのみ依存し決定する性質を持つ確率過程のモデルである。そのため、将来に渡って得られる最終的な累積報酬 R_t を式(1)のように表すことができる。

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (1)$$

ここで r_t は離散時間 t における実報酬、 γ は割引率である。割引率とは、遠い将来に得られる報酬ほど小さく評価するものであり、 $0 \leq \gamma \leq 1$ を満たす。そしてこの累積報酬 R_t を用いて、状態 s の価値 $V^\pi(s)$ は式(2)のように定式化できる。

$$V^\pi(s) = E_\pi\{R_t | s_t = s\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\right\} \quad (2)$$

ここで s_t は離散時間 t における状態とし、 π は方策を示す。 $E_\pi\{\cdot\}$ はエージェントが方策 π に従う時の期待値を示す。つまり状態 s の価値 $V^\pi(s)$ は方策 π に依存するため、各状態 s における価値 $V^\pi(s)$ を最大化する方策を探すことが学習の目的である。

本システムの問題ではユーザ周辺の照度と、ユーザがどの要求をしているかという2種類の離散的な状態(ユーザ周辺照度は照度という連続値を一定の範囲で区切ることで離散的と捉える)が存在する。行動は照度変化量の決定という連続的なものである。そのため、本システムでは強化学習手法の中でも連続値を行動として扱うことに適しているActor-Criticアルゴリズムを用いる[8]。

3.2 Actor-Criticアルゴリズム

Actor-CriticアルゴリズムにおけるエージェントはActorとCriticにより構成される。Actorは保持する確率の方策 $\pi(s)$ (行動を一意にではなく確率的に選択する)に従って行動 a を決定する。Criticは各状態の評価を示す状態価値関数 $V(s)$ を持ち、Actorの実行した行動を評価して方策を更新する。Actor-Criticアルゴリズムの概念図を図2に示す。

図2 Actor-Criticアルゴリズムの概念図

以下に、図2の各ステップを説明する。

- (1) 状態観測
エージェントが環境から状態 s_t を受け取る。
- (2) 行動
Actorが確率の方策 $\pi(s_t)$ に従い、環境に対して行動 a_t を実行する。
- (3) 報酬
行動 a_t によって状態が s_{t+1} に遷移し、その評価としてCriticに報酬 r_t が与えられる。
- (4) 強化信号
Criticは報酬 r_t と次の状態 s_{t+1} を観測し、CriticとActorの学習指標となるTemporal Difference誤差(TD誤差)を計算する。TD誤差 δ_t は式(3)のように表される。

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \quad (3)$$

γ は割引率と呼ばれ、 $0 \leq \gamma \leq 1$ の係数である。このTD誤差 δ_t によってCriticは状態価値関数 $V(s_t)$ を、Actorは確率の方策 $\pi(s_t)$ を更新する。Criticの状態価値関数 $V(s_t)$ の更新は式(4)のように表される。

$$V(s_t) \leftarrow V(s_t) + \alpha_c \delta_t \quad (4)$$

α は学習率と呼ばれ、 $0 \leq \gamma \leq 1$ の係数である。ここでの α_c はCriticのための学習率であることを表す。またActorの確率の方策 $\pi(s_t)$ は、TD誤差 δ_t が正数であるなら a_t の選択確率が上がるように、負数であるなら選択確率が下がるようにパラメータを更新する。本システムでは確率の方策 $\pi(s)$ に正規分布を用いているため $\alpha_m \delta_t$ に従って平均を、 $\alpha_s \delta_t$ に従って標準偏差の値を更新している。

α_m はActorにおける平均のための学習率を表し,
 α_s はActorにおける標準偏差のための学習率を表す.

本システムの扱う状態はユーザ周辺照度と、ユーザ要求という2種類の離散的なものである。そのため従来のActor-Criticアルゴリズムでは、2種類の状態数を掛け合わせた数の状態を学習する必要がある(状態aの数が8個、状態bの数が4個の場合、 $8 \times 4 = 32$ 個の状態を学習しなければならない)[11, 12]。しかし2.2節で述べたように、本システムでは効率的な学習を行うことでユーザの操作負担を減らす必要がある。そこで効率の良い学習のために、ユーザ周辺照度に関するActorとユーザ要求に関するActorというように、2種類のActorを用いて2種類の状態を別々に学習するActor-Criticアルゴリズムを検討する。この手法であれば、2種類の状態数を足し合わせた数の状態を学習するだけで良い(状態aの数が8個、状態bの数が4個の場合、 $8 + 4 = 12$ 個の状態を学習するだけで良い)、学習効率の向上を期待できる。この2種類のActorを利用するアルゴリズムをTwo-Actor-Criticアルゴリズムと呼ぶこととする。

3.3 2種類のActorを利用するTwo-Actor-Criticアルゴリズム

Two-Actor-Criticはユーザ周辺照度に関するActorとユーザ要求に関するActorという2種類のActorを持つ。各要求において求められる照度変化量の周辺照度による増減率は各要求間で同じであるという仮定の下、ユーザ周辺照度に関するActorはユーザ要求に関するActorの増減率を決定するものとする。期待としてはユーザ周辺照度が低い程小さな値を、ユーザ周辺照度が高い程大きな値を選択することが望ましい。この2種類のActorの関係を図3に示す。なお、本システムではActorの確率的方策として正規分布を用いる。

a_t^i は離散時間 t でのユーザ周辺照度に関するActorの行動、 a_t^d はユーザ要求に関するActorの行動を表す。また s_t^i はユーザ周辺照度に関する状態(「0-1000 lx」「1000-2000 lx」など)、 s_t^d はユーザ要求に関する状態(「とても明るく」「やや暗く」など)を表す。図3に示すように、まずユーザ周辺照度に関するActorの確率的方策 $\pi^i(s_t^i)$ における行動 a_t^i を決定する。そして、ユーザ要求に関するActorの確率的方策 $\pi^d(s_t^d, a_t^i)$ は平均の基礎となる値と a_t^i を掛け合わせた値を正規分布の平均として行動を決定する。このように本システムでは2種類のActorを備えた学習アルゴリズムを用いることを検討する。このTwo-Actor-Criticアルゴリズム

図3 2種類のActorの関係

図4 Two-Actor-Criticアルゴリズムの概念図

の概念図を図4に示す。

図4に示すように、本アルゴリズムにおいてActorはユーザ要求とユーザ周辺照度に関する2種類のActorが存在するが、Criticはユーザ要求に関する1種類しか存在しない。なぜならCriticの役割はActorの行動を評価することであり、本システムにおいてはActorの行動によってユーザ要求が「満足」の状態にどれだけ近付いたのかが行動を評価する唯一の指標だからである。以下に、図4における各工程を説明する。

(1) 状態観測

エージェントが環境から状態 $[s_t^i, s_t^d]$ を受け取る。

(2) ユーザ周辺照度に関するActorの行動

ユーザ周辺照度に関するActorが確率的方策 $\pi^i(s_t^i)$ に従って行動 a_t^i を決定する。

(3) ユーザ要求に関するActorの行動

ユーザ要求に関するActorが確率的政策 $\pi^d(s_t^d, a_t^i)$ に従い、環境に対して行動 a_t^i を実行する。 $\pi^d(s_t^d, a_t^i)$ における a_t^i の働きは図3に示した通りである。

(4) 報酬

行動 a_t^i によって状態が $[s_{t+1}^i, s_{t+1}^d]$ に遷移し、その評価としてCriticに報酬 r_t が与えられる。

(5) 強化信号

Criticの状態価値関数 $V^d(s_t^d)$ と2種類のActorの確率的政策 $\pi^i(s_t^i)$ 、 $\pi^d(s_t^d, a_t^i)$ の更新は3.2節で述べたものと同様に、TD誤差 δ_t を用いたものである。ただし、確率的政策 $\pi^i(s_t^i)$ と $\pi^d(s_t^d, a_t^i)$ は値のスケールが異なるため、TD誤差を反映させる際の学習率を個別に設定する必要がある。

図5 仮想ユーザの感覚尺度の例

4. シミュレーション実験

本問題において、Two-Actor-CriticアルゴリズムがActor-Criticアルゴリズムと比べて効率の良い学習アルゴリズムであることを示すために実験を行った。

4.1 実験内容

本実験では、周辺照度に応じて各要求に必要な照度変化量が変わるという、人間の感覚尺度を模倣した仮想ユーザを設計し、それを各アルゴリズムで学習した。下記にシミュレーション上での、実験手順を示す。

- (1) 周辺照度を初期照度に設定する。
- (2) 仮想ユーザは自身の保持する目標照度を満たせるように要求を決定する。
- (3) 学習アルゴリズムは仮想ユーザからの要求に応じて周辺照度を変更する。
- (4) 周辺照度と目標照度との差が許容照度差以下になれば(2)に戻る。許容照度差以下になれば仮想ユーザが「満足」の状態になったとする。

上記の工程を1ステップとし、1つのシステムにつき指定のステップ数、シミュレーションを行った。

4.1.1 仮想ユーザの概要

仮想ユーザの感覚尺度の例を図5に示す。

図5の横軸はユーザ周辺照度を示す。縦軸は目標照度からユーザ周辺照度を引いた照度である。つまり図5はユーザ周辺照度および目標照度までの差によって、ユーザの要求(「とても明るく」や「ほんの少し明るく」など)が定まることを示す。例えば、ユーザ周辺照度

表1 仮想ユーザ1の感覚尺度

ユーザ 周辺照度	閾値1	閾値2	閾値3
0-1000 lx	+20 lx	+40 lx	+80 lx
1000-2000 lx	+80 lx	+160 lx	+320 lx
2000-3000 lx	+320 lx	+640 lx	+1280 lx
3000-∞ lx	+1280 lx	+2560 lx	+5120 lx

が500 lxで目標照度までの差が+1000 lxの時には仮想ユーザは「とても明るく」という要求をするが、ユーザ周辺照度が1500 lxで目標照度までの差が+1000 lxの時には「明るく」という要求をする。システム上で仮想ユーザが感覚尺度として実際に保持しているのは図5に示した各ユーザ要求の境目となる各閾値(lx)のデータである。本実験で使用した仮想ユーザ1の感覚尺度を表1に示す。ここでは図5に示すように「ほんの少し明るく」と「やや明るく」の境目を閾値1、「やや明るく」と「明るく」の境目を閾値2、「明るく」と「とても明るく」の境目を閾値3とする。

表1には「明るく」系統の要求しか記載していないが、目標照度からユーザ周辺照度を引いた照度がマイナスの場合には、「暗く」系統の要求を出すために表1の各閾値をマイナスにしたものを適用する。

4.1.2 仮想ユーザの検証

本問題では2種類の状態が存在するが、この仮想ユーザ1はActor-Criticアルゴリズムがその2種類の状態に対応しない場合とする場合で差が出ることを目的に設計されている。そのため、その差が無いのであれば本研究で改訂した感覚尺度の性質を模倣できてい

ないことを意味し、本実験の対象問題として相応しくない。ここで、2種類の状態に対応しない場合はユーザ要求だけで照度変化量を判断することを意味し、このActor-Criticアルゴリズムはユーザ要求に関する8個の状態しか学習しない。一方で対応する場合はユーザ要求とユーザ周辺照度の両方により照度変化量を判断することを意味し、ユーザ周辺照度が4個の状態に分かれているため、 4×8 の32個の状態を学習することになる。この2つのActor-Criticアルゴリズムが仮想ユーザ1を学習した結果に差が生じるかを検証するために予備実験を行った。その実験結果を図6に示す。また、本実験におけるActor-Criticアルゴリズムのパラメータを表2に示す。表2における割引率や学習率などは強化信号の大小に影響を与えるものであり、強化信号が大きい場合は粗雑且つ高速に学習を行うことを、小さい場合は慎重且つ低速に学習を行うことを意味する。本対象問題ではより速く学習を進めることが求められる。そのため正しく学習できる範囲で、最も速く学習が行えるパラメータを経験則に基づいて設定した。

ここで、図6の入力数について説明する。本システムでは、ユーザが部屋に入る度に要求照度に到達できるようにシステムに希望を指示するが、希望照度に到達するまでに複数回の要求が必要である。図6の入力

図6 1種類および2種類の状態に対応したActor-Criticアルゴリズムの各ステップでの入力数

表2 Actor-Criticアルゴリズムのパラメータ

ステップ数	200回
割引率	0.5
Criticの学習率	0.5
Actorの学習率	0.1
「満足」状態遷移時の報酬	800
その他の状態遷移時の報酬	-50
目標照度	ランダム [0 lx~4000 lx]
初期照度	ランダム [0 lx~4000 lx]
許容照度差	100 lx

数とは1ステップの間にユーザが要求を入力し直した複数回の要求回数を指す。部屋へ入る回数がステップ数と対応しており、ステップ数が進展するごとにシステムは学習を行うので入力数が減少すると考えられる。また、図6の点は各ステップにおける30試行の入力数の平均値である。そして、実線は2種類の状態に対応したActor-Criticアルゴリズムの入力数を、破線は1種類の状態に対応したActor-Criticアルゴリズムの入力数を対数近似したものである。

図6に示すように2つの対数近似曲線を見ると、2種類の状態に対応したActor-Criticアルゴリズムの方が入力数が少なくなっている。つまり、仮想ユーザ1は2種類の状態に対応した方が良い学習が行える設計になっていることが分かる。

4.2 結果

Two-Actor-CriticアルゴリズムとActor-Criticアルゴリズムで、仮想ユーザ1を学習した際の入力数の推移(30試行の平均値)を図7に示す。ここでのActor-Criticアルゴリズムとは図6における2種類の状態に対応したものである。本実験におけるTwo-Actor-Criticアルゴリズムのパラメータを表3に示す。本実験でも表2のパラメータ設定と同様に、学習速度と学習精度のバランスを考慮したパラメータを経験則に基づいて設定した。また、目標照度、初期照度、許容照度差、ステップ数の設定およびActor-Criticアルゴリ

図7 仮想ユーザ1でのTwo-Actor-CriticアルゴリズムとActor-Criticアルゴリズムの各ステップでの入力数

表3 Two-Actor-Criticアルゴリズムのパラメータ

割引率	0.5
Criticの学習率	0.5
Actor (ユーザ要求) の学習率	0.1
Actor (ユーザ周辺照度) の学習率	0.0005
「満足」状態遷移時の報酬	300
その他の状態遷移時の報酬	-50

ズムのパラメータは4.1.2項の実験時と同じである。

図7の横軸はステップ数、縦軸は入力数を示す。また図7の点は各ステップにおける30試行の入力数の平均値である。そして、実線はTwo-Actor-Criticアルゴリズムの入力数を、破線はActor-Criticアルゴリズムの入力数を対数近似したものである。図7に示すように実線で表すTwo-Actor-Criticアルゴリズムは破線で表したActor-Criticアルゴリズムと比べて、学習精度は同等且つ学習速度が速いという結果になった。

4.3 考察

本実験においてはユーザ要求に関する状態が8個、ユーザ周辺照度に関する状態が4個であり、Actor-Criticアルゴリズムは 4×8 の32個の正規分布を最適化しなければならない。一方でTwo-Actor-Criticアルゴリズムでは $4 + 8$ の12個の正規分布を最適化するだけで良い。そのため、Two-Actor-CriticアルゴリズムはActor-Criticアルゴリズムと比べて学習速度が速いという実験結果が得られたと考えられる。この考察を検証するために、更にユーザ周辺照度に関する状態数を増やした仮想ユーザ2を用いて実験を行った。仮想ユーザ2の感覚尺度を表4に示す。

表4 仮想ユーザ2の感覚尺度

ユーザ 周辺照度 \ 閾値	閾値1	閾値2	閾値3
0-500 lx	+10 lx	+20 lx	+40 lx
500-1000 lx	+20 lx	+40 lx	+80 lx
1000-1500 lx	+40 lx	+80 lx	+160 lx
1500-2000 lx	+80 lx	+160 lx	+320 lx
2000-2500 lx	+160 lx	+320 lx	+640 lx
2500-3000 lx	+320 lx	+640 lx	+1280 lx
3000-3500 lx	+640 lx	+1280 lx	+2560 lx
3500- ∞ lx	+1280 lx	+2560 lx	+5120 lx

図8 仮想ユーザ2でのTwo-Actor-CriticアルゴリズムとActor-Criticアルゴリズムの各ステップでの入力数

表4に示すように、仮想ユーザ2はユーザ周辺照度に関する状態が8個に分かれている。仮想ユーザ2を4.1.2項の実験と同様にTwo-Actor-CriticアルゴリズムとActor-Criticアルゴリズムで学習した際の入力数の推移(30試行の平均値)を図8に示す。なお、各アルゴリズムで使用したパラメータは4.1.2項の実験時と同じである。

図8の横軸はステップ数、縦軸は入力数を示す。また図8の点は各ステップにおける30試行の入力数の平均値である。そして、実線はTwo-Actor-Criticアルゴリズムの入力数を、破線はActor-Criticアルゴリズムの入力数を対数近似したものである。

図7では200ステップの時点における、Two-Actor-CriticアルゴリズムとActor-Criticアルゴリズムの対数近似曲線の差は図7の結果よりも図8の結果の方が大きくなった。このことから上述した考察が正しく、Two-Actor-Criticアルゴリズムは2種類の状態を2種類のActorで対応することで学習効率を向上できていると言える。

5. 実環境における実験

Two-Actor-Criticアルゴリズムが、ユーザの感覚尺度を学習し得るのかを検証するために人を対象とした実験を行った。

5.1 実験環境

本実験は図9に示す実験室にて行った。

図9に示す実験室の大きさは縦6.15m、横8.26m、高さ2.75mであり、天井には48灯の蛍光灯を備えている。これらの蛍光灯は個別に光度を272cd～1090cdの間で調節可能である。次に本実験システムの操作画面を図10に示す。

図10に示すように、本実験ではコンピュータの操作画面に「明るく」「やや明るく」「ほんの少し明るく」「ほんの少し暗く」「やや暗く」「暗く」「満足」の7つのボタンを配置した。「とても明るく」「とても暗く」のボタンを排除した理由は、本実験室の照明が実現できる照度幅では一般ユーザがそれらの要求を出す場面が少ないことが予備実験から判明したためである。また「満足」ボタンの右側には目標照度とユーザ周辺照度との差を表示した。これは純粋な学習の効果を抽出するため、実験におけるヒューマンエラーを極力回避するためである。

本実験では被験者に学習システムと無学習システムの2種類の照明制御システムを操作させた。学習システムとはTwo-Actor-Criticアルゴリズムを用いてユーザの感覚尺度を学習するものである。ただし、学習の

図9 実験室

図10 操作画面

際に更新する要素が多ければ多い程時間が必要となる。そのため、本実験では被験者の負担を考慮し、更新する要素を2種類のActorの平均のみとした。つまり、各Actorの標準偏差と状態価値関数の値は予備実験から導いた適切な値を予めから設定しておき、更新は一切行わないものとした。一方、無学習システムとは上記の学習システムから、あらゆるパラメータの更新を無くしたものである。

5.2 実験内容

被験者は20歳代の男女10名(男性9名、女性1名)であった。実験の手順を以下に示す。なお、下記目標照度と初期照度はシステムに予め設定されているものである。

- (1) 被験者に可能な限り少ない要求入力回数で「満足」の状態に辿り着けるように教示を与える。
- (2) 実験システムは被験者の周辺照度が目標照度になるように照明を点灯させる。

- (3) 被験者は目標照度の明るさを覚える。
- (4) 実験システムは被験者の周辺照度が初期照度になるように照明を点灯させる。
- (5) 被験者は目標照度になるように要求を入力する。
- (6) 実験システムは被験者からの要求に応じて照明の明るさを変化させる。
- (7) 周辺照度と目標照度との差が許容照度差以下になれば(4)に戻る。許容照度差以下になれば「満足」の状態になったとする。

工程(3)に「被験者は目標照度の明るさを覚える。」とあるが、人の視覚は光環境の変化に順応するため、正確に目標照度を記憶し続けることは難しい。本システムにおいて、視覚の順応によるヒューマンエラーが「満足」の判定に影響することは学習に支障を来す。そのため本実験では工程(7)に示すように、「満足」の判定は被験者ではなくシステムが決めるものとした。

上記の工程を1ステップとし、被験者には1つのシステムにつき20ステップ操作させた。そして、1つのシステムで20ステップが終わると、同様にもう1つのシステムで20ステップ操作させた。このシステムの操作順は被験者間でカウンタバランスを取った。また、ユーザ周辺照度に関する状態は「0-500 lx」「500-1000 lx」「1000-1500 lx」「1500-2000 lx」「2000-∞ lx」の5個に区切った。

5.3 結果

Two-Actor-Criticアルゴリズムを用いた学習システムと無学習システムを被験者が操作した際の入力数の推移(10人の平均値)を図11に示す。また、Two-Actor-Criticアルゴリズムのパラメータを表5に示す。本実験でも表2のパラメータ設定と同様に、学習速度と学習精度のバランスを考慮したパラメータを経験則に基づいて設定した。

図11 被害者が学習システムと無学習システムを操作した際の各ステップでの入力数

表5 実環境における実験でのTwo-Actor-Criticアルゴリズムのパラメータ

ステップ数	20 回
割引率	0.8
Actor (ユーザ要求) の学習率	0.8
Actor (ユーザ周辺照度) の学習率	0.0001
「満足」状態遷移時の報酬	600
その他の状態遷移時の報酬	-300
目標照度	1200 lx
初期照度	2700 lx
許容照度差	100 lx

図11の横軸はステップ数、縦軸は入力数を示す。また図11の点は各ステップにおける10人の入力数の平均値である。そして、実線は学習システムの入力数を、破線は無学習システムの入力数を線形近似したものである。図11が示すように、学習システムの方が無学習システムと比べて入力数が減少するという結果となった。

5.4 考察

5.3節に示すように、学習システムと無学習システムの入力数の推移に差が出る結果となった。これらのシステムの違いはTwo-Actor-Criticアルゴリズムに従った2種類のActorの平均更新の有無である。つまり、無学習システムに比べて学習システムの入力数が減少した要因はTwo-Actor-Criticアルゴリズムによる学習ということになる。このことから、Two-Actor-Criticアルゴリズムはユーザの明るさに対する感覚尺度を学習し得ると考えられる。

6. 今後の展望

5章の実験では被験者の負担を考慮したため、学習を簡略化して実験を行った。そのため、今後各Actorの標準偏差と状態価値関数の更新も含めた学習システムで実験を行う必要がある。また、今回は全ステップにおける目標照度と初期照度が同じであったため、これらをステップ毎に変更しても学習が成功するかを検証しなければならない。

そして今回の実験で示したのはユーザの感覚尺度を学習し得るということのみであり、人を対象とした場合にも提案アルゴリズムがActor-Criticアルゴリズムよりも効率的な学習が可能かを示す必要がある。その点における提案アルゴリズムの有効性を実環境での実験によって検証することが今後の課題として挙げられる。

更に、今回のシステムはユーザが1人であることを想定した。リビングルームやオフィスルームといった

実システムの環境下では複数人のユーザが存在し、それらの感覚尺度が異なることが考えられる。そのため、各ユーザの望む明るさが競合した際にどのような対応を取るべきかを今後検討しなければならない。

7. まとめ

本稿では照明制御システムにおける操作負担の増加という課題に対して、2種類のActorを利用するアルゴリズムを用いた感覚的操作による照明制御システムを提案した。提案アルゴリズムは本対象問題に存在する2種類の状態に対して2種類のActorを用いることで、より効率的な感覚尺度の学習を目的としたものである。

本稿ではこのアルゴリズムの概要を説明すると共に、有効性を示すための実験について述べた。一つは仮想ユーザを用いたものであり、提案アルゴリズムはActor-Criticアルゴリズムよりも高い学習効率を持つことを示した。もう一つは実環境における実験であり、提案アルゴリズムが人の明るさに対する感覚尺度を学習し得ることを示した。

今回の実環境における実験は簡略化したものであったため、今後は異なる実験環境や全ての学習パラメータの更新を含んだ学習システムでの実験を通し、このアルゴリズムが人の感覚尺度をどの程度の精度や速度で学習可能であるかを検証する。更に、人を対象とした場合にも提案アルゴリズムが既存のActor-Criticアルゴリズムよりも高い学習効率を持つかを検証することが課題である。

参 考 文 献

- [1] V Singhvi, A Krause, C Guestrin, JH Garrett Jr, HS Matthews : Intelligent light control using sensor networks, Proceedings of the 3rd international conference on Embedded networked sensor systems, pp.218-229, 2005.
- [2] B Brumitt, JJ Cadiz : "Let There Be Light" Examining Interfaces for Homes of the Future, Proceedings of Interact '01, pp.375-382, 2001.
- [3] K Gajos, DS Weld : SUPPLE - Automatically Generating User Interfaces-, Proceedings of the 9th international conference on Intelligent user interfaces, pp.93-100, 2004.
- [4] M Ashibe, M Miki, T Hiroyasu, : Distributed Optimization Algorithm for Lighting Color Control using Chroma Sensors, 2008 IEEE International Conference on Systems Man and Cybernetics, pp.174-178, 2008.
- [5] LC Throop : Field of play - sensual interface -, Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp.82-86, 2003.

- [6] K Tsukada, M Yasumura : Ubi- Finger – Gesture Input Device for Mobile Use –, Proceedings of APCHI 2002, Vol.1, pp.388–400, 2002.
- [7] SS Stevens : On the psychophysical law, Psychological Review, Vol.64(3), pp.153–181, 1957.
- [8] AG Barto, RS Sutton, CW Anderson : Neuronlike adaptive elements that can solve difficult learning control problems, Neurocomputing; foundations of research, pp.535–549, 1988.
- [9] T Shikakura, H Morikawa, Y Nakamura : Perception of Lighting Fluctuation in Office Lighting Environment, Journal of Light and Visual Environment, Vol.27, No.2, pp.75–82, 2003.
- [10] DH Kelly : Visual Responses to Time-Dependent Stimuli. I. Amplitude Sensitivity Measurements, Journal of the Optical Society of America, pp.422–429, 1961.
- [11] J Morimoto, K Doya : Acquisition of Stand- up Behavior by a Real Robot using Hierarchical Reinforcement Learning, Proceedings of the Seventeenth International Conference on Machine Learning, pp.623–630, 2000.
- [12] GR Gajjar, SA Khaparde, P Nagaraju, SA So-man : Application of actor-critic learning algorithm for optimal bidding problem of a GenCo, IEEE Transactions on Power Engineering Review, Vol.18, No.1, pp.11–18, 2003.
- [13] M Miki, T Hiroyasu, K Imazato, M Yonezawa : Intelligent Lighting Control using Correlation Coefficient between Luminance and Illuminance, Proc IASTED Intelligent Systems and Control, Vol.497, No.078, pp.31–36, 2005.
- [14] M Miki, E Asayama, T Hiroyasu : Intelligent Lighting System using Visible-Light Communication Technology, 2006 IEEE Conference on Cybernetics and Intelligent Systems, pp.1–6, 2006.
- [15] T Mitchell, B Buchanan, G DeJong, T Dietterich, P Rosenbloom, A Waibel : Machine Learning, Annual Review of Computer Science, Vol.4, No.1, pp.417–433, 1990.
- [16] RS Sutton, A Barto : Reinforcement Learning – An Introduction –, The MIT Press, 1998.

(2010年12月15日 受付)

(2011年 5 月 8 日 採録)

[問い合わせ先]

〒610-0321 京都府京田辺市多々羅都谷1-3

同志社大学

廣安 知之

TEL : 0774-65-6932

FAX : 0774-65-6932

E-mail : tomo@is.doshisha.ac.jp

著者紹介

ひろやす ともゆき
廣安 知之 [非会員]

1997年早稲田大学大学院理工学研究科後期博士課程修了。早稲田大学理工学部助手を経て、1998年同志社大学工学部助手。2003年より同志社大学工学部助教授。2007年より同志社大学生命医科学部教授。進化的計算、最適設計、並列処理、設計工学等の研究に従事。IEEE、電気情報通信学会、計測自動制御学会、日本機械学会、超並列計算研究会、日本計算工学会各会員。医療情報システム研究室/知的システムデザイン研究室。1997年早稲田大学大学院理工学研究科後期博士課程修了。遺伝的アルゴリズムを始めとする最適化手法の並列化および知的化の研究を進めている。

なかむら あきゆき
中村 彰之 [非会員]

2011年同志社大学大学院工学研究科修士課程修了。

みき みつのり
三木 光範 [非会員]

現在：同志社大学大学院情報工学専攻教授。1950年生。1978年大阪市立大学大学院工学研究科博士課程修了、工学博士。大阪市立工業研究所研究員、金沢工業大学助教授を経て1987年大阪府立大学工学部航空宇宙工学科助教授。1994年同志社大学工学部教授。研究分野はシステム工学、最適化、並列処理など。最近は並列処理と最適化を組み合わせた技術をオフィス照明の分野に展開し、知的照明システムを研究・開発している。著書は「工学問題を解決する適応化・知能化・最適化法」(技法堂出版)、「進化する人工物」(オーム社)、「理系発想の文章術」(講談社現代新書)、「頭が良くなる知的生産の技術」(中経出版)等多数。

よし み まさと
吉見 真聡 [非会員]

2004年慶應義塾大学理工学部情報工学科卒業。2009年同大学院理工学研究科開放環境科学専攻後期博士課程修了。博士(工学)。2006年度より日本学術振興会特別研究員(DC1)。現在、同志社大学理工学部助教。リコンフィギャラブルシステム、並列処理、知的システムの研究に従事。電子情報通信学会、人工知能学会各会員。

よこうち ひきたけ
横内 久猛 [非会員]

1971年徳島大学工学部電気工学科卒。同年(株)日立製作所入社。中央研究所にて眼底写真の自動認識の研究、がん細胞・血液像の自動認識装置の開発、カラーブラウン管検査自動化の研究、電子線描画露光装置の開発、デジタルX線装置の開発などに従事。1991年(株)日立メディコに転属。実時間デジタルX線装置開発、医療画像診断装置のマーケティング担当。2006年(株)日立メディコ退職。2008年同志社大学生命医科学部教授。医用画像工学・信号処理、ヒト脳機能計測・処理などの研究に従事。日本生体医工学会、日本医用画像工学会、映像情報メディア学会、日本ヒト脳機能マッピング学会各会員。

The Sensuous Lighting Control System using Actor - Critic Algorithm

by

Tomoyuki HIROYASU, Akiyuki NAKAMURA, Mitsunori MIKI, Masato YOSHIMI
and Hisatake YOKOUCHI

Abstract :

In this research, we create the lighting system that users can order brightness sensuously such as “very lighten” or “little darken” by using Actor - Critic algorithm. This system needs to learn the sensory scale so that it can act depending on 2 kinds of statuses, which are the demand of a user and the brightness around the user. We built up Actor-Critic algorithm with 2 kinds of Actors applying to 2 kinds of states. In this paper, the overview and effectivity of this algorithm are shown.

Keywords : Lighting, Sensory scale, Actor - Critic algorithm

Contact Address : **Tomoyuki HIROYASU**

Doshisha Universit

1-3 Tataramiyakodani Kyotanabe, 610-0321, JAPAN

TEL : 0774-65-6932

FAX : 0774-65-6932

E-mail : tomo@is.doshisha.ac.jp