

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 -

同志社大院 富田 浩司 同志社大工 三木 光範 同志社大工 廣安 知之

Application of Reinforcement Learning to Intelligent Network Systems

- Construction of Intelligent Lighting Systems by Q-Learning -

Koji TOMITA, Graduate School of Engineering, Doshisha University

Mitsunori MIKI, Faculty of Engineering, Doshisha University

Tomoyuki HIROYASU, Faculty of Engineering, Doshisha University

Abstract: In this paper, we apply Q-Learning to intelligent network systems. The intelligent network system is the autonomous and distributed systems and it is constructed with the intelligent artifacts. The intelligent artifacts have three factors: sense, judge and act. Each intelligent artifact of the intelligent network system has the same purpose and it tries to satisfy the purpose by itself. In this paper, Q-Learning is applied to judge part of the intelligent artifacts. Through the numerical examples, the effectiveness of the proposed system and Q-Learning is made clarified.

1 はじめに

近年、電子デバイス技術と情報処理技術の驚異的な発展により、電化製品、自動車、航空機をはじめとする多くの人工物は自身の制御や管理を使用者や環境に合わせて自律的に行うことで利用者(人間)の負担を軽減する、いわゆる「知的化」が行われている。そこで、我々はそれらの賢さを持つ機器を知的人工物と定義し、その知的性が生まれるメカニズムの解析し、より知的な機器・システムの開発を行っている^[1]。一方、最近のシステムはインターネットの普及に伴って、家庭内機器をネットワーク化するホームネットワーク、人・道路・車両をネットワーク化する ITS など、ネットワーク化が急速に進展している。この「知的化」、「ネットワーク化」の2つの観点から、我々は知的な人工物をネットワーク化した自律分散型のネットワークシステムを提案しており、これを「知的ネットワークシステム」と呼んでいる^[2]。本システムでは、ネットワークに与えられて目的を各機器が自律的に取り込み、ネットワーク全体でその目的を満たすため、機器の故障による機能低下などのトラブルが起こった場合でも柔軟に対処し、ユーザの満足を維持し続けることができる。しかし、大きな課題の1つに各機器の自律制御の最適化問題がある。

本論文では、我々が提案している知的ネットワークシステムの各機器の制御に強化学習を適用することによって、従来の手法よりも効率よく目的を達成することを示す。具

体的には、強化学習の代表的な手法である Q-Learning を用いた知的照明システムのシミュレーションによって、その有効性を検証する。

2 工学的人工物と知的人工物^[1]

2.1 人工物の定義と分類

人為的に作られた、いわゆる人工物には、建物、機械、自動車、航空機、家電製品、通信網などで代表される工学的人工物と言語、知識、規則、法律、組織などで代表される社会的人工物、さらに小説、絵画、彫刻などの芸術的人工物がある。

工学的人工物は物質を基にして作り出される形ある「もの」であり、社会的人工物は物質とは無関係な、形の無い「もの」である。これらの人工物に共通の属性は目的である。すなわち、これらの人工物は人間が何らかの目的を持って作り出したものであり、その人工物の機能や性能の評価はこの目的に沿って行われる、いわば道具としての人工物である。芸術的人工物は表現媒体として何かの物質を利用しており、形のある「もの」であるが、作り出す側も、それを利用する側も明確な目的を意識せず、制作者は自己表現として、利用者は鑑賞するものとして捉え、人工物の機能や性能の評価を行うことができない独特な人工物である。

一方、これらの人工物の他に、品種改良した農作物やバイオテクノロジーを利用して生み出された自然界には

存在しない自然物がある。これは人為的に作られたという意味では人工物であるが、内部のメカニズムは人工的ではなく自然物であり、いわば人工的自然物といえる。

このように、人工物を大きく分けると以下の4つに分類できる。

1. 工学的人工物：建物，自動車，家電製品など
2. 社会的人工物：言語，規則，法律など
3. 芸術的人工物：小説，絵画，彫刻など
4. その他の人工物：品種改良した農作物など

我々が第1段階として取り組まなければならないのは工学的人工物である。なぜなら、工学的人工物は電子デバイス技術と情報処理技術により実現されるものが多く、明確な目的を持ち、その機能や性能が評価できるからである。

2.2 知的人工物

著者らの一人は、上記で述べた人工物のうち明確な目的を持ち、その機能や性能が評価できる工学的人工物に限定し、基本的な考察を行っている。その考察で、工学的人工物の中には、人工物自身の制御や管理を環境に合わせて自律的に行うことで、利用者（人間）の負担を軽減するような知的性質を持つ工学的人工物が多く存在することが述べられており、それを「知的人工物」と呼んでいる。

知的人工物の定義は「人工物が、使われる環境や利用者の仕方に依存する多くのパラメータを持ち、これらの組み合わせにより、多様な利用者要望や使用環境に柔軟に対応できるように設計されている時、センスした情報と与えられた知識や学習で得た知識を基に、適切な組み合わせを人工物自身が選択し、利用者の要望や環境に応じて最高の機能と性能を提供してくれる時、その人工物を知的人工物と呼ぶ」としている。

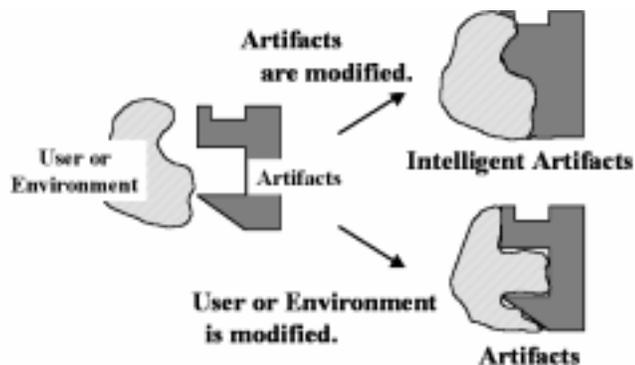


Fig. 1 Interfaces of artifacts and user/environment.

Fig.1は人工物とそれを取り巻く人間環境と自然環境の関係を模式的に表したものである。知的でない人工物で

は、人工物側のインタフェースが変化せず、人間や自然環境に負荷が作用し、人工物の性能が十分に発揮されない。一方、知的人工物では人工物側のインタフェースが変化し、人間環境や自然環境に負荷が少なく、人工物の性能が十分に発揮される。

以上から、工学的人工物での「知的」という概念は、利用者の感覚として人工物が人間の知性と呼ばれるに近いような性質を具備している場合に使われ、知的人工物は人間が行っていた人工物の運用と管理をすることだと考えられる。具体的な例としては、マイコン制御の電気釜（人間に代わり火加減を調節する）、ニューロ洗濯機（汚れの度合いを自動検出する）、人間の入退出や窓側の明るさに応じ電灯のON/OFFを制御しているような知的照明、利用形態に応じ柔軟に通信網構成を変えることのできるインテリジェントネットワークなどがある。すなわち、知的人工物は人工物の管理・運用の自動化能力という、これまでの人工物とは異なった特徴を持つ。一般的には知能や知的という言葉は人間が持つような高度な知能を考えるが、ここではどんなに単純でもここで定義した性質を持つものは知的人工物である。例えば、パイメタルで構成される単純な温度調節器も1つの知能であり、それを備えた人工物はそれによって高度な運用・管理ができていれば知能を有していると考えられる。

2.3 知的人工物の知的構造

知的人工物とは2.2節に述べたように、最近よく見られる知的な人工物の総称であるが、この定義に従うと、知的人工物は利用者を含む広義の環境条件の変化に対応して人工物自身のパラメータを自律的に変化させるために、その環境条件の変化をセンスするための各種のセンサが必要である。次に、センサで得た情報を基に判断し、それに沿って人工物のパラメータを変化させる動作ができなくてはならない。すなわち、全ての知的人工物は知的性質としてこの3つの要素を持ち、Fig.2で表すことができると考えられる。

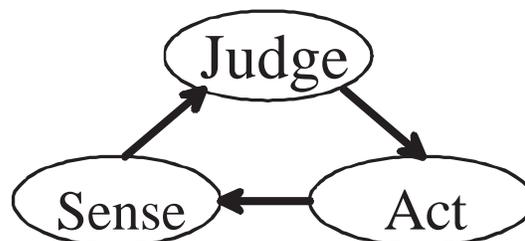


Fig. 2 Factors of intelligent artifacts.

例えば、VTRには多くの知的機能が備わっており、ジャストクロック機能は、毎日同じ時間に自動的に電源が入り、NHKの時報を検知し、内蔵時計の進み遅れを判断し、そのズレを補正し電源を切る。光感知照明機器は、外の

明かりをセンスし、あらかじめ組み込まれている明るさの判断基準から、光束を調節する知的人工物である。現在では知的とは言えない自動ドアも、人をセンスし、人の有無の判断基準から、ドアの開閉を制御するため、知的人工物の一つであると考えられる。

3 知的ネットワークシステム

3.1 知的ネットワークシステムの概要

知的ネットワークシステムとは先に述べた知的人工物を複数ネットワーク化したものであり、本システムの特徴は Fig.3 に示すように、主制御器を持たず、部屋・建物等のネットワーク全体に対してユーザが要求する「目的」を与えるだけで、後は接続された各知的人工物が知的性質を使って自律的に最適に動作してくれることである。

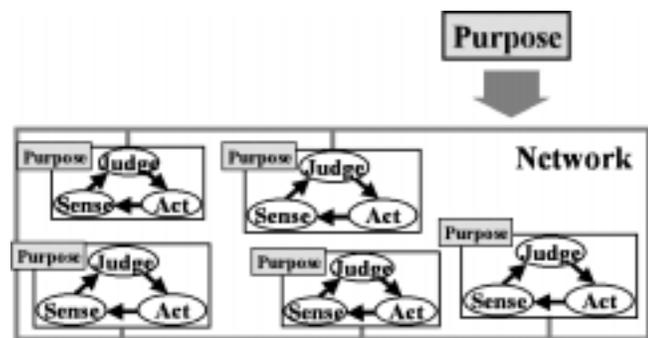


Fig. 3 Conceptual diagram of intelligent network systems.

各知的人工物は既存の予め組み込まれている目的があるが、その達成すべきあらゆる目的を外部からネットワークに流すことによって、各知的人工物をその方向へ向かわせる。具体的には、知的人工物はネットワークに接続されると、各自が同じ目的を取り込み、その目的に合った判断基準を自ら生成する。そして、知的人工物が個々に持つ各種センサからセンスされた情報と生成された判断基準を基にそれぞれが自律的に動作する。実際は、各自が目的を満たすよう動作するだけであり、他の知的人工物と協調するのではないが、結果として、ネットワーク全体としてより知的に動作しているように見える。これにより、ネットワークに接続された知的人工物の数のみで目的を満たすように動作できると考えられる。

本システムの有効性は次のようになる。

1. ネットワークに「目的」を与えておくだけでよい。
2. 機器のネットワークへの参入・離脱が容易である。
3. 1つの機器では不可能な作業を行える。
4. ある機器の故障時に起こる機能低下を他機器によって柔軟に対応し、補うことができる。
5. 既存機器のみで新しい機能を生み出せる。

6. ネットワークの機能を有していれば、システムに接続する知的人工物の種類は問わない。

本システム的具体例としては、ある建物において「部屋を快適にしろ」という目的を与えると、接続されている知的照明、知的エアコンなどが「部屋の温度を28度に維持し、人がいる所だけを明るくする」等の明るさの判断基準・温度の判断基準を自ら生成する。各知的人工物は部屋の温度が28度になるように、また、人がいる所だけが明るくなるように動作し、部屋を快適にする。ある機器が故障した場合も、他の知的人工物により対処することが可能となる。また、交通システムにおいては、多くの交通機器（知的人工物）をネットワークに接続しておくことにより、例えば「交通渋滞を防げ」という目的を与えておくと、各交通機器はユーザからの命令を待たず、信号機故障や交通事故時による交通渋滞を解消するように自律的にネットワーク内で対処することが可能となる。

3.2 知的ネットワークシステムの課題

知的ネットワークシステムの課題はいくつかあるが、最も大きな課題の1つに各機器の自律制御の最適化問題がある。各知的人工物が目的に合った判断基準を自律生成し、各種センサからの外部情報のみからその判断基準を基に目的を満たすよう制御させるのは極めて困難である。また、目的をどこからどのようなデータ形式で与えるのか、各知的人工物の動作情報や位置情報をどのようにネットワーク全体に送るのか、などのプロトコル問題もある。

4 強化学習を用いた知的照明システム

4.1 知的照明システムの概要

知的照明システムは知的ネットワークシステムの一つであり、本システムの基礎的な検討を行うのに使用している。システム構成は Fig.4 に示すように、複数の知的照明機器（以下 知的照明）をネットワークに接続し、ネットワークに与えた「人がいる所を $X [lx]$ の明るさにせよ。」という共通の目的を満たせるかどうかを検証するものである。

ここで用いられる知的照明は人感知センサと明るさ感知センサの両方が備わっているものとし、各知的照明は、各真下の人の有無と明るさ $[lx]$ をセンスし、人の有無に合わせて調光することができるタイプを用いた。各知的照明の調光パターンは光度 $0 \sim 1,000 [cd]$ である。

4.2 従来の知的照明システム^[2]の問題点

従来の手法では2.2節の、ネットワークに「目的」を与えておくだけでよいこと、機器のネットワークへの参入・離脱が容易であること、1つの機器では不可能な作業を行えること、ある機器の故障時に起こる機能低下を他機器によって柔軟に対応し補うことができることの有効

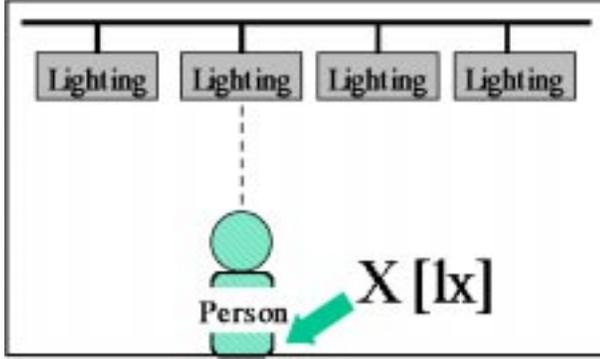


Fig. 4 Intelligent Lighting Systems

性は検証されているが、最も大きな課題である各機器の自律制御の最適化問題はあまり検討されていない。

知的照明システムの従来の自律制御アルゴリズムを以下に示す。

- (1) 各知的照明は一齐に、各々一度だけランダムに動作してみる ($\pm 20[\text{cd}]$)。
- (2) 人の真上にいる知的照明は、(1) 後の環境 (人がいる場所の照度) をセンスし、その情報をネットワーク全体に送る。
- (3) 各知的照明は (2) の行動によって、目的への達成度が上がったかどうかを判断する。上がったならば、各知的照明はもう一段階上の動作を行う。下がったならば、再度 (1) の動作を行う。
- (4) この手順の繰り返しにより、他の情報、自分の動作の有効性がわからなくても、知的照明全体で目的を満たすように動作することができる。

従来の手法では、目的に合った判断基準を各知的人工物が自律的に生成するのではなく、動作前後での目的への達成度の比較という判断基準を予め各知的人工物に持たせ制御しており、ある程度の成果はできるが柔軟性はなくこれ以上の機能は期待できないという問題点がある。そこで、各機器の制御に強化学習のような目的に合った判断基準を自律生成する手法の適用が必要となる。

4.3 Q-learning を用いた知的照明システム

強化学習^[3]とは移動などの行為の行うエージェントが教師付き学習 (Supervised learning) のような直接の教師を持たずに、行為に対する環境からの報酬のみから、適切な行為の学習を行う自律的学習である。

強化学習で最も代表的なアルゴリズムに Q-learning^[4]がある。Q-learning では、エージェントは状態認識器、行動選択器と学習器の 3 構成要素からなる。状態認識器は、状態と行動の対のテーブルすなわちルールベースで、各ルールは Q 値と呼ばれる重みを持っている。行動選択器には Boltzmann 選択、 ϵ -greedy 選択などがあるが、次

式に示すように $\exp(Q(s,a)/T)$ に比例した割合で行動を選択する Boltzmann 選択が広く用いられている。

$$p(a_i|x) = \frac{e^{Q(x,a_i)/T}}{\sum_{k \in \text{actions}} e^{Q(x,a_i)/T}} \quad (1)$$

学習器では次式に従って Q 値を更新する。ここで、 α は学習率、 γ は割引率である。

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a_t)] \quad (2)$$

あるスケジュールに従って学習率 α を減少させ、多数の試行の後に Q 値が収束すると、各状態における最大の Q 値を持つルールの選択が最適な政策となることはすでに証明されている。

この Q-learning を知的照明システムの自律制御に適用した。これにより、ネットワーク化された各機器の目的に合った判断基準を自動生成が可能となり、さらに学習による目的達成時間の効率化が期待できる。

Q-learning を用いた知的照明システムのアルゴリズムを次に示す。

- (1) 人の真上にいる知的照明は現在の環境 (人がいる場所の照度) 状態 S を観測し、他へ送る。
- (2) 各知的照明はある行動選択方法 (1) に従って光束を強めるか弱めるかを定める。
- (3) 各知的照明は報酬 r を受け取る。
- (4) 人の真上にいる知的照明は次の環境 (人がいる場所の照度) 状態 S を観測し、他へ送る。
- (5) 各知的照明はそれらの情報を基に (2) 式により Q 値を更新する。
- (6) この手順を繰り返す。

4.4 シミュレーション

シミュレーションは 2 種類行った。シミュレーション 1 では、Q-learning を用いた知的照明システムにおいて、1 つの知的照明では不可能な明るさを「目的」とした場合に、各知的照明が協力して目的を満たせるかを検証するものである。シミュレーション 2 では、ある知的照明が故障した場合による機能低下を残りの知的照明によって柔軟に対応し、与えられた目的を満たし続けられるかを検証するものである。

4.4.1 シミュレーション 1

シミュレーション 1 では、Q-learning を用いた知的照明システムにおいて、1 つの知的照明では不可能な明るさを「目的」とした場合に、各知的照明が協力して目的を満たせるかを従来の手法と比較する。シミュレーション画面を Fig.5 に示し、各パラメータ設定を Table1 に示す。目的

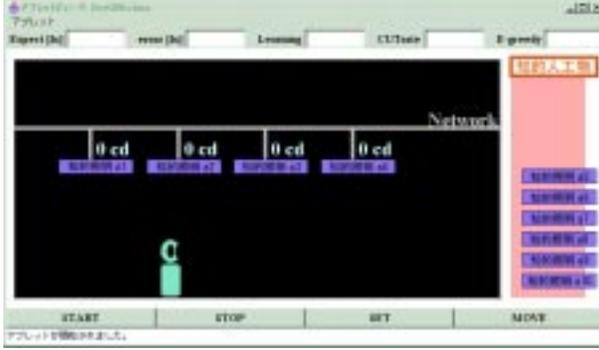


Fig. 5 Simulation of Intelligent Lighting Systems

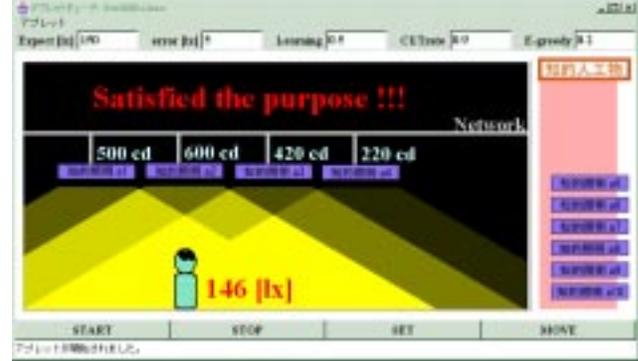


Fig. 6 Result of Intelligent Lighting Systems by Q-learning.

Table 1 Parameters

目的照度	150[lx]
誤差	± 5 [lx]
状態 S の数	60 状態 (S0 ~ S59)
行動 A の数	2 状態 (A0, A1)
各 Q 値の初期状態	0.1(全て共通)
学習率	=0.5
割引率	=0.9
行動選択方法	ボルツマン選択 (T=0.2)
報酬	ゴール rwd=100, 他 0

照度を 150[lx] としたのは、1 台の知的照明では 100[lx] が限界であるからである。状態 S は 5[lx] 単位で 60 状態に分割し、S0(0 ~ 5[lx]) ~ S59(195 ~ 300[lx]) とし、各状態 S における行動 A は 2 状態、A0(+20[cd])、A1(-20[cd]) とした。学習率、割引率、温度定数等は、予備実験による経験的な知見を参考に設定した。報酬は目的を達成したときに rwd=100 をその他の状態では rwd=0 を与え、ネットワーク化した知的照明は 4 台とした。

シミュレーション 1 における、Q-learning を用いた知的照明システムの結果の例を Fig.6 に示す。様々な調光パターンがあるが、必ず目的を満たせることが確認できた。Fig.7 には Q-learning を用いた知的照明システム、Fig.8 には比較として従来の知的照明システムの目的を達成するまでの軌跡を示す。横軸は目的を達成するまでにかかった時間であり、縦軸は合計照度である。各知的照明が消えている状態から目的が達成されるまでを 1 試行とし、表には 1 試行目、10 試行目、30 試行目を示した。Fig.8 については従来の制御手法には学習機能がないため、1 試行目、10 試行目、30 試行目の意味は特にない。

Fig.8 からわかるように、従来の知的照明システムは設計者が予め与えておいた制御で動作し、学習機能も無いため、目的を満たすまでの時間が 10 試行目において 50 ステップかからないが、30 試行目には 150 ステップかかり、

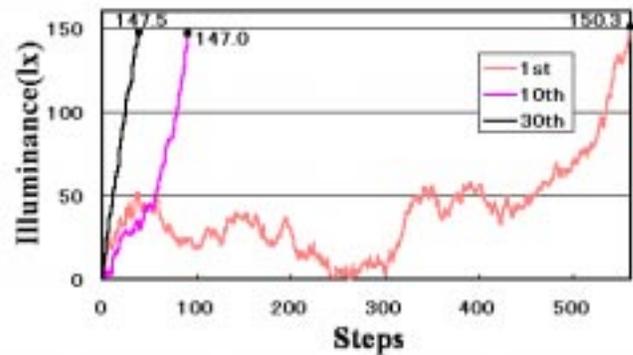


Fig. 7 Steps and the illuminance of intelligent lighting systems by Q-learning (simulation 1).

試行回数に関わらず時間が不安定である。一方、Fig.7 からわかるように、Q-learning を用いた知的照明システムでは試行回数を重ねるほど学習していき、1 試行目はまだ学習がなされていないため 550 ステップもかかっているが、10 試行目になるとある程度学習により判断基準ができ、30 試行目には目的を達成するのに 50 ステップかかっていない。ここでは表示していないが、この後何試行繰り返しても 30 試行目と同じ制御を行い、完全な判断基準が生成されているといえる。これにより、Q-learning を用いた知的照明システムは従来の手法よりも効率よく目的を満たすことができることがわかり、さらに Q 値のルールベースという判断基準が確立しさえすれば、後はそれを使って最適に制御することができるため、判断基準を予め組み込んでおく必要がない。これは、知的ネットワークシステムにおいて極めて大きな成果といえる。

4.4.2 シミュレーション 2

シミュレーション 2 では、ある知的照明が故障した場合による機能低下を残りの知的照明によって柔軟に対応し、与えられた目的を満たし続けられるかを確認する。各パラメータはシミュレーション 1 と同じである。ここでは、シミュレーション 1 の Fig.6 の状態から左から 3 台目の知

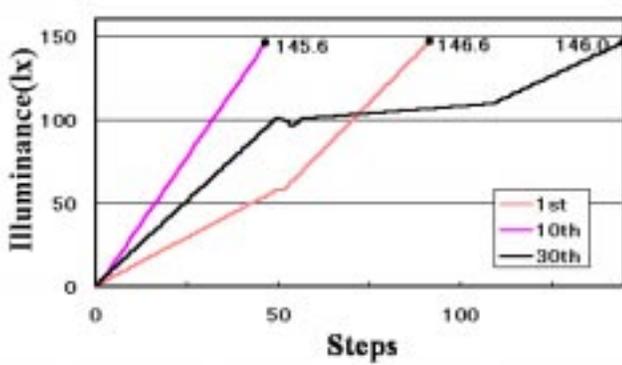


Fig. 8 Steps and the illuminance of the conventional intelligent lighting systems (simulation 1).

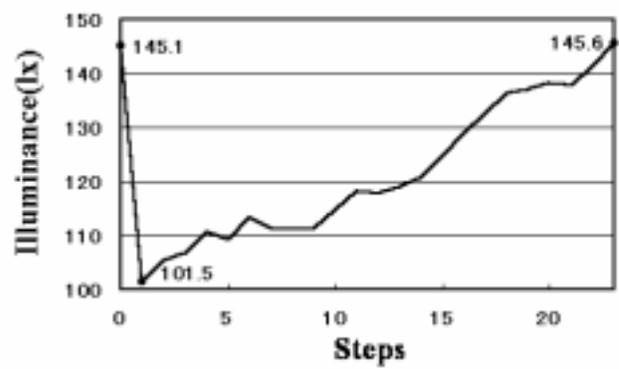


Fig. 10 Steps and illuminance (simulation 2).

的照明を壊したときのシミュレーション結果を Fig.9 に、ある知的照明が故障してから、残りの機器が光束を強め目的照度達成するまでの軌跡を Fig.10 に示す。横軸は目的を達成するまでにかかった時間であり、縦軸は残りの3台での合計照度である。

を用いることにより、従来の知的照明システムでは実現できなかった各機器の判断基準の自動生成を行うことができた。また、学習を用いることで、従来の手法よりも目的を達成までの時間が短縮され、効率化が行えた。ある機器の故障時における他機器の柔軟な対応もシミュレーションによって検証できた。

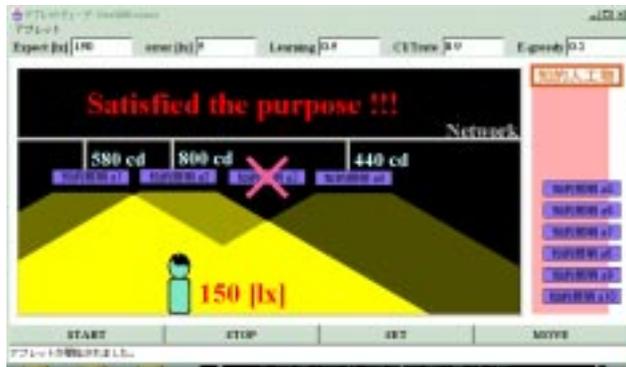


Fig. 9 Result of Intelligent Lighting Systems (simulation 2).

これまでの研究によって、目的照度が変われば新しい判断基準が生成されることがわかった。また、同じ目的照度が同じでも、多くの環境で学習を行うと新しい判断基準が多数生成されることもわかった。実際に本システムを使用する際、ユーザからの目的や使用環境は日々変化すると考えられるが、その都度、各知的照明が学習によって判断基準を獲得するのではあまり有効なシステムとはいえない。今後の課題としては、生成された判断基準を目的ごとに一カ所にデータベース化し、各知的照明が与えられた目的と今の環境に適した最適な判断基準を学習を通して選択し、自身に取り込むことができれば、より有効なシステムになると考えられる。

Fig.9 および Fig.10 からわかるように、ある知的照明が故障した時点 (Steps=0) では人のいる場所の明るさは一度約 100[lx] にまで下がるが、すぐ (Steps=22) に残りの知的照明が光度を強め、結果として残りの3台で人がいる場所に 150[lx] の目的を維持するように動作していることがわかる。

謝辞

本研究は文部省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「知能情報科学とその応用」における研究の一環として行った。

ここでは検討していないが、2台の知的照明を壊したとき、残りの2台で目的を達成することができない場合がある。その場合でも新しい知的照明を唯ネットワークに接続すれば、後は自律的に目的を満たせることが検証されているが、実現不可能な目的が与えられてた際のシステムの対応については今後の課題である。

参考文献

5 結論と今後の課題

本論文では、知的ネットワークシステムへの強化学習の適用を行った。具体的には、知的ネットワークシステムの基礎的検討に使用している知的照明システムにおいて各機器の自律制御に代表的な強化学習である Q-learning

- [1] M.Miki and T.Kawaoka : Design of Intelligent Artifacts: A Fundamental Aspects , Proc.JSME International Symposium on Optimization and Innovative Design(OPID97), 1997-9
- [2] 廣安, 三木, 富田: 知的人工物を用いた次世代ネットワークシステム~知的照明システムによる基礎的検討~, 日本機械学会第9回設計工学・システム部門講演会, pp.518-521 (1999)
- [3] 畝見: 強化学習, 人工知能学会誌, pp.830-836(1994)
- [4] Watkins, C.J.C.H, and, Dayan, P: Technical Note: Q-Learning, R.S.Sutton(ed.), Reinforcement Learning, pp.55-68, Kluwer Academic(1993)