ユーザ教示とデータ通信によるQ-table生成機構を有するユビキタスQ学習エンジンの実装

岡田量太[†,☆] 田川聖治[††] 塚本昌彦[†††]

近年，ユビキタスコンピューティング環境を実現するために，様々な小型デバイスが開発されている．しかし，既存の小型デバイスは，環境やユーザの好みの変化への適応が十分には行えない．本論文では，学習機能を有する小型デバイスとして，ユビキタスQ学習エンジンを提案する．はじめに，ユビキタスQ学習エンジンの要件について述べる．次に，代表的な強化学習アルゴリズムであるQ学習に，ユーザ教示機能や，ネットワーク通信によるQ-tableの交換機能などを付加した新たな学習アルゴリズムを示す．さらに，小型のマイクロコントローラを用いたユビキタスQ学習エンジンの設計と実装について述べる．最後に，ユビキタスQ学習エンジンを，空調服のファンの速度制御に応用し，その有用性を実験によって検証する．

Implementation of Ubiquitous Q-learning Engine with Q-table Generation Mechanism by User's Instructions and Data Communication

Ryota Okada,[†,☆] Kiyoharu Tagawa[††] and Masahiko Tsukamoto[†††]

Recently, various small devices have been developed for realizing ubiquitous computing environments. However, conventional small devices are not sufficiently support adaptation of their functions to the changes of environments and users' tastes. In this paper, a ubiquitous Q-learning engine is proposed as a small device that has a learning function. First of all, the requirements for a ubiquitous Q-learning engine are described. Then, a new learning algorithm is proposed for the ubiquitous Q-learning engine. The proposed learning algorithm is a revised version of the Q-learning, i.e., a typical reinforcement-learning algorithm, added some unique functions, namely the user's instruction function, the Q-table's exchange function with the network communication, and so on. Furthermore, we show our design and implementation of the ubiquitous Q-learning engine by using a microcontroller. Finally, the ubiquitous Q-learning engine is applied to the speed control of an air-conditioning fan fixed in clothes, and the usefulness of the ubiquitous Q-learning engine is verified through several experiments.

1. はじめに

近年，コンピュータの小型化，高性能化，無線通信技術の発達にともなって，「誰でも・いつでも・どこででも」コンピュータリソースにアクセスすることができるユビキタスコンピューティング（ubiquitous computing）環境[1),2)]の実現に向けた様々なシステムや機器の研究開発が行われている．

これまでユビキタスコンピューティングのための小型デバイスとして，温度，加速度，磁気，バイオなどの10種類のセンサを有する無線センサネットワーク構築用ノードのMOTE[TM 3)]，日用品に装着することを目的としたSmart-Its[4)]，イベント駆動型ルール処理エンジンのAhroD[TM 5)]などが提案されている．

しかし，これらのデバイスでは，その挙動の多くの部分がデバイスへのプログラミングの段階で固定されてしまうため，想定される様々な状況に応じた機能を事前にプログラミングしておくか，ルールなどの形式で動作パターンを設定しておく必要があり，設計者への負担が大きい．また，実環境にはつねに多くの不確

† 神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University
†† 近畿大学理工学部
School of Science and Engineering, Kinki University
††† 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University
☆ 現在，株式会社日立製作所
Presently with Hitachi Co. Ltd.

実性が存在し，設計者が想定していない状況に遭遇した場合や，ユーザの嗜好や使用目的が変化した場合など，ルールやプログラムの入れ替えが必要となる可能性がある．

本論文では，ユビキタスコンピューティングのための小型デバイスに，現場でのパラメータチューニングによるオンラインの適応能力を付加することを目的とする．そのため，学習機能を有する小型デバイスを考え，ユビキタス Q 学習エンジン（以降，Q 学習エンジン）[6] と名付ける．Q 学習エンジンでは，強化学習の 1 つである Q 学習[7] にユーザ教示機能を付加した独自の学習アルゴリズムを採用する．教示機能によってユーザの意図や好みを素早く Q 値に反映したり，パソコンやほかの Q 学習エンジンと Q-table を交換する通信機能を備える．さらに，Q 学習エンジンでは，ユーザの衣服に装着して使用することを想定したいくつかの設計上の工夫も行った．

2. 関 連 研 究

報酬を最大化する行動則を学習する強化学習は，動的計画法との関連が明らかにされたことを契機に理論的な研究が進んでいる[7]．また，強化学習を中心とする様々な学習アルゴリズムが，ロボットによるサッカーゲーム（RoboCup）[8] や，人工市場での取引（U-Mart）[9] など，複雑で動的な問題に適用され，その有用性が実証されている．さらに，小型のコンピュータに学習アルゴリズムを搭載した例として，RoboCup の実機ロボットリーグ[10] がある．しかし，RoboCup や U-Mart は，仮想的な世界におけるゲームにすぎず，実世界の動的な環境への適応が考慮されていない．また，実機ロボットリーグのロボットも，サッカー競技を目的とした専用マシンであり，ユビキタス環境において人々の生活を支援するための汎用性を持たない．

人間を学習システムの中に組み込んだ例としては，産業用ロボットに対する作業の教示があり，オフラインプログラミングとオンラインプログラミングに大別される．また，オンラインプログラミングは，ティーチングプレイバック方式とも呼ばれ，ロボットの作業環境の中で，教示者がロボットを作業手順どおりに動かしながら，ロボットの動作データとしてプログラムを作成する方法である[11]．このため，ユーザの好みや環境の変化にオンラインで適応することができないという問題がある．一方，教示者からロボットが適切な行為として教示情報を受け取り，行動ルールを獲得する学習の手法として，対話的クラシファイアシステムが提案されている[12]．しかし，遠隔操作によるロボットの教示学習では，ロボットとユーザが同じ空間内に存在しないため，誤った学習による被害をユーザが直接受けることはない．このため，ロボットに教示を行うユーザの安全性や不安感などは考慮されておらず，ロボットに対する教示学習では，外界からロボットを眺める教示者の視点と，ロボットの視点との違いが大きな問題となる[12]．

ユビキタスコンピューティング環境の 1 つとして，ホームネットワークによる情報家電の制御がある[13]．情報家電を学習によって知的に制御しようという取り組みは数多く行われており，Q 学習を利用したものとしては，ネットワーク化された照明システムの制御がある[14]．これらはユビキタス環境で学習を導入することの有用性を示すものとして，本研究のアプローチを支持するものと見なすことができる．ただし，これらの研究では，処理能力が高いホームサーバや，屋内に敷設されたネットワークを想定している．これはリッチでクローズドなコンピュータ環境を前提とするものであり，コンピュータ化が進んでいない環境や屋外などのインフラを前提としないような場所や，ウェアラブルコンピューティングのような計算リソースに制限がある場合にも，様々な情報機器をうまくパラメータチューニングして連係動作させたいような状況を考えると，これらの方式の適用は困難となる．本論文では，そのようなインフラを前提としないような環境を考え，学習アルゴリズムの選択において，プログラムの処理時間やデータの容量など，実装上の制約を受けることの少ない小型デバイス上での学習アルゴリズムを考えた．

3. ユビキタス学習エンジン

3.1 学習エンジンの必要性

多くの人々は知能を持った機械に興味や期待をいだきながらも，同時に，知能を持った他者の存在に不安や恐怖を感じる．このため，強化学習に関する研究は，長い歴史を持ち，いくつもの成果をあげながらも，ホーム・ロボットの自律的な行動は制限され，学習機能を持つ家電製品はほとんど見られない．一方，Web の検索エンジンなどソフトウェアの世界では，ユーザの行動パターンを学習する機能など，すでに常識となっている．あらゆるモノや場所に遍在するコンピュータの恩恵を享受するためには，知能を持った他者との平和的な共存の道を探る必要がある．前田ら[15] も「妖精・妖怪」と呼ばれる高い知性を持った機械と共存する 50 年後のユビキタス社会を予想している．今回開発した Q 学習エンジンは，上記のような成熟したユ

ビキタス社会の実現に向けた取り組みの一環である．

ここで，Q 学習エンジンなど学習機能を有する小型デバイス（以降，学習エンジン）を用いることで，学習機能のない既存の小型デバイスと比較して，以下のような利点が考えられる．

- 不確実性のある現実環境において，設計段階で想定していなかった状況に陥った場合でも，ユーザがその場で望む動作を教示することにより，システムのチューニングが行える．
- 小型デバイスの行動パターンの生成が部分的に自動化されるため，設計者の負担を大幅に軽減することができる．
- 試行錯誤の繰返しによる機械的な学習によれば，ユーザが考えるよりも優れた行動パターンを偶然に発見する可能性がある．

3.2 学習エンジンの要件

ユビキタスコンピューティング環境で用いられる学習エンジンには以下の要件が求められる．

ユビキタス性

学習エンジンは様々な場所やモノに取り付けられる．したがって，有線や無線による通信機能を持ち，複数の学習エンジン間で相互にデータを交換することができ，多種多様なセンサやアクチュエータを取り付けることが可能で，小型で廉価であることが求められる．

学習の効率性

学習エンジンの実際の環境での利用を考えた場合，学習の効率性が大きな問題となる．特に Q 学習のような強化学習は，試行錯誤を繰り返すことにより環境に適応していく枠組みであるために，ユーザが小型デバイスを利用しながら，オンラインで学習を行うことは非常に効率が悪い．このため，学習効率の良い新たな学習アルゴリズムが求められる．

安全・安心性

学習エンジンは人間の生活に密着した空間に備え付けて動作させることを前提としている．このため，ユーザに危害を及ぼすような出力を行ったり，ユーザに不安感を与えたりする恐れがある．これらの問題を解消するため，学習エンジンにはユーザ側のヒューマン・エラーも考慮した安全設計が求められる[16]．

3.3 学習エンジンの設計方針

Q 学習エンジンの設計において，上記の 3 つの学習エンジンの要件を満たすための設計方針を示す．

まず,「ユビキタス性」の要件を満たすため，Q 学習エンジンには通信機能を付加するとともに，その実装においては小型化に努める．このため，超小型のマイクロコンピュータを採用するとともに，電子部品の選択やレイアウトを工夫する．また，スイッチの個数についても極力減らし，複数のスイッチの組合せにより Q 学習エンジンを操作できるようにする．

次に,「学習の効率性」については，Q 学習の収束性を高めるため，ユーザの動作を教示として Q 値に反映させる新たなオンライン学習アルゴリズムを考案する．また，通信機能を積極的に利用して，オフライン学習や，ほかの Q 学習エンジンで得られた Q-table を取得し，Q 学習の際の初期値として使用する．

最後に「安全・安心性」については，Q 学習エンジンとユーザで適切なコミュニケーションがとれるような工夫を行う．すなわち，Q 学習エンジンの内部状態をボード上の LED などでリアルタイムに表示させるとともに，パソコンとの通信によって Q 学習エンジンが獲得した Q-table の内容を観察したり，変更したりできる開発アプリケーションを作成する．

また，ユーザの安全を確保するため，上記のオンラインの学習アルゴリズムでは，Q-table に基づき Q 学習エンジンが選択した行動と，ユーザが指示した行動が異なる場合，つねにユーザの指示を優先して実行する．さらに，ユーザの思いどおりに Q 学習エンジンが動作しない場合を想定し，Q 学習エンジンのデバイス出力の制御を停止して，ユーザが直接マニュアルで出力を操作できる機能を付ける．

そのほか，ユーザの操作ミスを減らすためには，アプリケーションに適したアフォーダンスを有するインタフェースを採用する必要がある[16]．そこで，Q 学習エンジンには様々なタイプのユーザ教示用スイッチが接続できるような入力端子を準備する．

以下に，上記の設計方針に基づき考案した学習アルゴリズムや，ハードウェアの実装について述べる．

4. 学習アルゴリズム

強化学習における最も代表的な学習アルゴリズムに Q 学習があり，実際のロボットの制御にも採用され，その有効性が実証されている[17]．そこで，小型デバイスのための学習アルゴリズムは，既存の Q 学習をベースとし，効率的なオンライン Q 学習を実現するため，ユーザ教示による Q 値の更新機能と，ネットワーク通信を利用した妥当な Q 値の獲得機能を付加した．

4.1 Q 学習アルゴリズム

まず，通常の Q 学習[7] について説明する．Q 学習ではセンサ入力から決定された状態 s_i（$i = 0$〜m）に対して，各行動 a_j（$j = 1$〜n）の良さを Q 値と呼ばれる実数値 $Q_{ij} = Q(s_i, a_j)$ で表す．また，現在の状態に最も適した行動を Q 値に基づき決定する．さ

らに，これらの Q 値を記録するのが Q-table であり，Q 学習とは Q-table を適切に更新することであると考えられる．通常の Q 学習のアルゴリズムを以下に示す．α は学習率，γ は割引率，R は報酬である．

(1) センサ入力から状態 s_t を判定する．
(2) Q 値に基づき，行動 a_t を決定する．
(3) 行動 a_t を実行する．
(4) センサ入力から次状態 s_{t+1} を判定する．
(5) 環境から報酬 R を受け取る．
(6) Q 値に基づき，行動 a_{t+1} を決定する．
(7) 式 (1) により Q 値を更新する．

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

(8) $s_t \leftarrow s_{t+1}$，$a_t \leftarrow a_{t+1}$
(9) 手順 (3) へ戻る．

4.2 ユーザ教示による Q-table の更新

通常の Q 学習においては，環境から得られる報酬と過去の Q 値に基づき Q-table の内容が更新される．一方，提案するユビキタス Q 学習では，ユーザが適切な行動を Q 学習エンジンに直接入力し，Q 値に反映させることで，Q 学習の効率性を高めている．すなわち，Q 学習エンジンが Q 値に基づき決定した行動と，ユーザが教示した行動が一致すれば，Q 値にボーナス B を与え，一致しなければ Q 値にペナルティ P を与える．また，ユーザが入力した行動を優先して実行することで，ユーザの安全を確保している．

提案するユビキタス Q 学習により，Q-table を更新する手順を，以下の手順 (1)〜(13) に示す．通常の Q 学習のアルゴリズムに対し，新たにユーザ教示機能のための手順 (4)〜(7) を追加している．

(1) センサ入力から状態 s_t を判定する．
(2) Q 値に基づき，行動 a_t を決定する．
(3) 行動 a_t を実行する．
(4) ユーザ教示入力 b_t を検知して，教示入力 b_t がなければ手順 (8) へ．あれば手順 (5) へ．
(5) ユーザ教示入力 b_t とマシンが決定した行動 a_t が一致し，$b_t = a_t$ ならば手順 (8) へ．一致しなければ手順 (6) へ．
(6) ユーザ教示入力 b_t と合致する Q 値に式 (2) のようにボーナス B（$B > 1$）を与え，ユーザが好まない行動 a_t に対応する Q 値に式 (3) のようにペナルティ P（$0 < P < 1$）を与える．

$$Q(s_t, b_t) \leftarrow B \times Q(s_t, b_t) \quad (2)$$
$$Q(s_t, a_t) \leftarrow P \times Q(s_t, a_t) \quad (3)$$

(7) $a_t \leftarrow b_t$ と変更し，実行する．
(8) センサ入力から次状態 s_{t+1} を判定する．
(9) 環境から報酬 R を受け取る．
(10) Q 値に基づき，行動 a_{t+1} を決定する．
(11) 式 (4) により Q 値を更新する．

$$Q(s_t, a_t) \leftarrow (1-\alpha)\, Q(s_t, a_t) + \alpha[R + \gamma \max_a Q(s_{t+1}, a_{t+1})] \quad (4)$$

(12) $s_t \leftarrow s_{t+1}$，$a_t \leftarrow a_{t+1}$
(13) 手順 (3) へ戻る．

4.3 通信による Q-table の獲得

Q 学習エンジンでは，通信機能を利用して，ほかの Q 学習エンジンやパソコンから Q-table をダウンロードして利用できる．通信による Q-table の獲得は，使用目的や接続条件が同じ Q 学習エンジン間でなければ効果はなく，ユーザが異なれば理想とする Q-table の内容もまったく同じではない．しかし，ある程度妥当な Q 値を初期値とすることで，前述のユーザ教示による Q 学習の収束時間が短縮されることが期待できる．また，ユーザ教示による Q 学習では，ユーザ自身が体験していない状況下での動作の Q 値は変更されない．そこで，通信機能を利用することで，ほかのユーザの Q 学習エンジンから未体験の状況下で学習された Q 値を取得することができる．

以下に，ネットワーク通信によって Q 学習エンジンに Q-table をダウンロードする手順を示す．ただし，Q 学習エンジンにはメインの Q-table とサブの Q-table が存在し，ユーザ教示による Q 学習で Q 値が変更されるのは，メインの Q-table のみとする．

(1) メイン Q-table の内容（Q 値）を，サブ Q-table にコピーして保存する．
(2) ネットワーク通信で得られた Q-table の内容（Q 値）で，メイン Q-table を上書きする．
(3) Q 学習エンジンを使用してみることで，ユーザが新たなメイン Q-table を評価する．
(4) ユーザが満足できなければ，保存していたサブ Q-table の内容をメイン Q-table に戻す．

基本的にダウンロードの機能は，ユーザがダウンロードしたのちに Q 学習エンジンの動作が良いかどうかを適切に判断できる場合に有効である．

5. Q 学習エンジンの実装

5.1 ハードウェア構成

今回試作した Q 学習エンジンは，Microchip Technology 社製のマイクロコントローラである PIC（Peripheral Interface Controller）をコアとして使用している．このマイコンの動作速度は最大 50 MHz で速

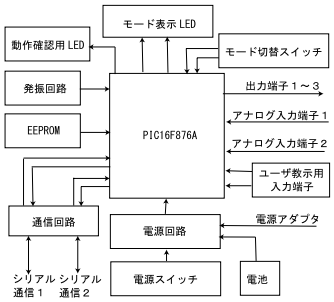

図 1 Q 学習エンジンの回路構成
Fig. 1 Structure of Q-learning engine.

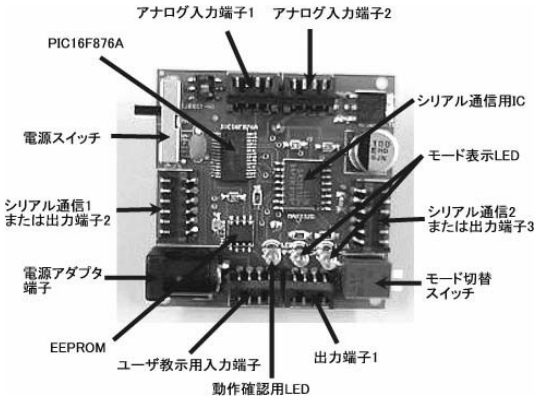

図 2 Q 学習エンジンの回路基板（サイズ：約 5 cm×5 cm）
Fig. 2 Circuit board of Q-learning engine (size: approx. 5 cm×5 cm).

くはないが，EEPROM やタイマ，AD 変換器などの内蔵モジュールを有し，消費電力も小さいため，近年様々な分野で使用されている．ただし，種類によって異なるがプログラムメモリは 512 ワード（1 ワード =14 ビット）から 16 k ワード，内蔵の EEPROM のサイズが 64 バイトから 256 バイトと記憶容量に制限がある．そこで，Q-table のサイズを考慮して外付けの EEPROM（32 k バイト）を使用し，マイコンにはプログラムメモリが 4 k ワードと十分な PIC16F876A を採用した．Q 学習エンジンの回路構成を**図 1**，回路基板を**図 2** に示す．

まず，図 1 に示した Q 学習エンジンの入出力端子の種類と個数は以下のとおりである．

- モード切替用スイッチ（1 ビット × 2 個）
- ユーザ教示用入力（1 ビット × 2 個）
- センサ入力：アナログ入力（10 ビット × 2 ポート）
- 出力：デジタル出力（2 ビット × 3 ポート）
- 通信：シリアル通信（2 ポート）
- モード表示 LED（2 個）
- マイコン動作確認用 LED（1 個）

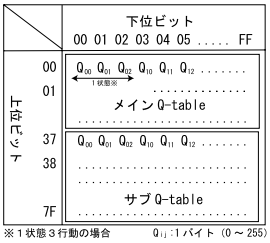

図 3 Q-table のデータ構造
Fig. 3 Data structure of Q-table.

- 電源（5 V）：ボタン電池，電源アダプタ端子

センサ入力は，温度センサ，湿度センサなど様々なアナログセンサに対応している．また，2 ポートのシリアル通信により，ほかの Q 学習エンジンとパソコン，または，最大 2 個までの Q 学習エンジンと同時に通信を行うことができる．モード切替用スイッチとユーザ入力端子の使用方法については後述する．さらに，Q 学習エンジンの電源は，電源スイッチにより，ボタン電池と電源アダプタを切り替えることができる．家電製品などに組み込む場合は電源アダプタから電源を供給し，衣服などに装着して利用する場合はボタン電池を使用することを想定している．

5.2 Q-table のデータ構成

前述のように，Q 学習エンジンの Q-table はメイン Q-table とサブ Q-table があり，外付けの EEPROM 内に存在する．外付け EEPROM は 32 k バイトの容量があり，メイン Q-table はアドレス 0x0000 から 0x36FF まで，サブ Q-table は 0x3700 から 0x7FFF までそれぞれ約 14,000 バイトの容量を持つ．Q 学習エンジンでは 1 つの Q 値を 1 バイトで表すため，状態 s_i（$i = 0$〜m），行動 a_j（$j = 0$〜n）とすると，Q 値の総数は $m \times n$ 個となり，メイン Q-table，サブ Q-table ともに $m \times n \leq 14{,}000$ バイトを満たす範囲内で行動数，状態数を設定する必要がある．**図 3** に外付け EEPROM 内のアドレスとデータの対応関係を示す．ここで，Q 学習エンジンが外付け EEPROM の Q 値を読み込む場合，たとえば，状態 s_2，行動 a_1 の Q 値のアドレスは，2（状態番号）×n（行動数）+1（行動番号）のように計算できる．

5.3 運転モードと機能

新たに開発した Q 学習エンジンは，以下に述べるように 6 種類の運転モードを持つ．

学習モード

Q 学習エンジンのメインとなるモードである．4.2 節で示したユーザ教示によるユビキタス Q 学習アルゴリズムにより Q-table を更新していく．

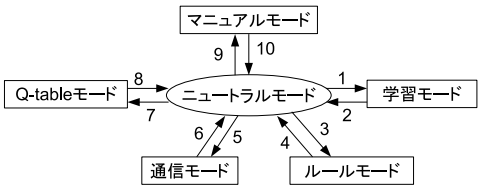

図 4 運転モードの遷移図
Fig. 4 Transition among operating modes.

ルールモード

学習モードで学習した Q-table や通信モードで得られた Q-table をもとに，センシングした状態 s_i に応じて適切な出力 a_j を決定し実行するモードである．ルールモードでは，学習が完了した状態を想定しており，4 章で述べた学習アルゴリズムによる Q-table の更新は行われない．このため，Q 学習エンジンは，既存のルール処理エンジン[5] と同様に機能する．

マニュアルモード

Q-table に基づく Q 学習エンジンの出力を無視して，ユーザからの入力をそのまま出力に伝えることで，Q 学習エンジンに取り付けられたアクチュエータをユーザが直接操作するモードである．ユーザの意思とは異なる行動パターンが学習された場合において，ユーザの危険を回避するための安全機能の 1 つである．

通信モード

Q 学習エンジンのシリアル通信機能を用いて，パソコン上のシミュレーションなどによりオフラインで学習した Q-table や，ほかの Q 学習エンジンで生成された Q-table，Q 学習のためのパラメータなどを Q 学習エンジンにダウンロードするモードである．また，自ら獲得した Q-table のコピーを，ほかの Q 学習エンジンやパソコンに送ることもできる．

Q-table モード

前述のように，Q 学習エンジンの EEPROM には，メイン Q-table とサブ Q-table が存在する．このモードでは，4.3 節で述べた通信による Q-table の獲得に際して，両者の Q-table の交換，および，メイン Q-table の初期化をユーザがスイッチ操作によって行う．

ニュートラルモード

任意のモードに移行するための中立的モードであり，学習やセンシングなどの機能は停止している．

5.4 スイッチの操作方法

運転モードの切替え

Q 学習エンジンには，前述のように 6 種類のモードがある．これらのモードの遷移図を**図 4** に示す．また，Q 学習エンジンのオンボード・スイッチによる図 4 の各運転モードの切替え方法を**表 1** に示す．図 4 の遷移図の枝番号は表 1 のモード遷移の番号に対応しており，1 ビット 2 個のモード切替スイッチを SW1 と SW2 とし，ユーザ教示用入力端子の 1 ビット 2 個のスイッチを UI1 と UI2 として，その操作を組み合わせることで運転モードを切り替えている．

表 1 運転モード切替え操作
Table 1 Switching operations for mode change.

モード	モード遷移	SW1	SW2	UI1	UI2
学習	1	ON	OFF	OFF	OFF
	2	OFF	OFF	OFF	OFF
ルール	3	OFF	ON	OFF	OFF
	4	OFF	OFF	OFF	OFF
通信	5	ON	ON	OFF	OFF
	6	OFF	OFF	OFF	OFF
Q-table	7	ON	ON	ON	ON
	8	OFF	OFF	OFF	OFF
マニュアル	9	OFF	OFF	ON	ON
	10	※ 1	※ 1	OFF	OFF

※ 1：SW1 か SW2 の一方を一度 ON にして OFF に戻す．

表 2 Q-table モードでの操作
Table 2 Switching operations in Q-table mode.

	UI1	UI2
Q-table 初期化	ON	ON
Q-table セーブ	ON	OFF
Q-table ロード	OFF	ON

表 3 通信モードでの操作
Table 3 Operations in communication mode.

Q-table 受信	UI1	UI2
シリアル通信 1	ON	OFF
シリアル通信 2	OFF	ON

Q-table の変換と初期化

Q-table の初期化，メイン Q-table からサブ Q-table への Q 値のセーブ，または，サブ Q-table からメイン Q-table への Q 値のロードは，Q-table モードにおいて，2 つのユーザ教示用入力端子（UI1，UI2）を**表 2** のように操作することで行う．

Q 学習エンジン間の Q-table 送受信

Q 学習エンジン間の Q-table の送受信は，両方の Q 学習エンジンが通信モードの状態で，データを受信する側の Q 学習エンジンからのユーザ操作による受信コマンドにより行われる．すなわち，ユーザ教示用入力端子（UI1，UI2）を**表 3** に示すように設定することで，それぞれシリアル通信 1 と通信 2 を区別し，送信側の Q 学習エンジンより Q-table をダウンロードする．また，その手順は以下のとおりである．

(1) 受信側 Q 学習エンジンのユーザ教示用入力端子 UI1，もしくは UI2 を ON にする．
(2) 現在使用中のメイン Q-table の内容を，表 2 のスイッチ操作でサブ Q-table に保存する．
(3) 送信側 Q 学習エンジンに対して送信リクエス

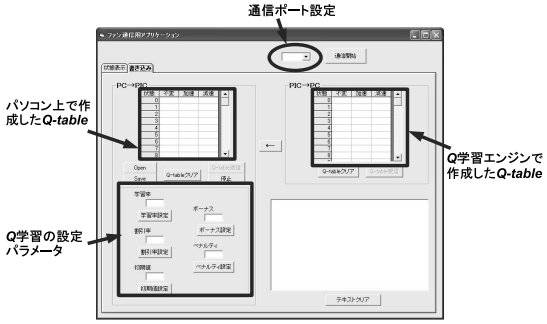

図 5 Q 学習エンジン開発アプリケーション環境
Fig. 5 Development application for Q-learning engine.

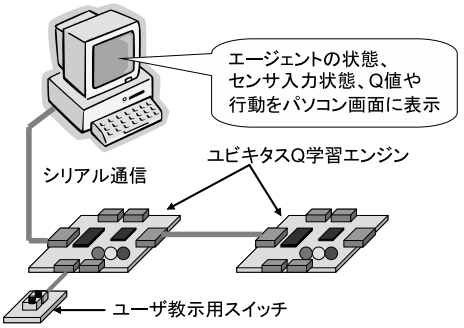

図 6 パソコン通信によるモニタリング
Fig. 6 Monitoring of Q-learning engine by PC.

ト信号を送り，Q-table を送信させる．

(4) 送信側 Q 学習エンジンより送られてきた Q-table の内容をメイン Q-table に保存する．

6. Q 学習エンジンの管理

パソコンによって Q 学習エンジンを管理するため，Q 学習エンジン開発アプリケーションを作成した．ここでは，Q 学習エンジン開発アプリケーションの 3 つの機能と操作方法について述べる．

6.1 Q-table の転送

シミュレーションや経験・知識に基づき，ユーザがパソコン上で作成した Q-table を，Q 学習エンジンにダウンロードすることができる．これにより，学習の手間を省けるほか，ある程度妥当な Q 値から学習を始めることで，教示を行うユーザの負担を軽減できる．ここで，Q 学習エンジンがパソコンから Q-table を受信する場合は，Q 学習エンジン開発アプリケーションを起動した状態で，Q 学習エンジン間での通信と同様，Q 学習エンジンが送信をリクエストするか，パソコン側で Q-table 送信ボタンを押す．

6.2 学習パラメータ設定

パソコンから Q 学習エンジンへのデータ送信は，前述の Q-table のほか，Q 学習のパラメータも送信して設定することができる．**図 5** のようなパソコン上の Q 学習エンジン開発アプリケーションにおいて，4.2 節で述べた学習率（α），割引率（γ），ボーナス（B），ペナルティ（P）の値をキーボードから入力する．ただし，学習率，割引率，ペナルティは 0〜1，ボーナスは 1 以上の範囲とする．次に，マウスで送信ボタンをクリックすることで Q 学習エンジンにパラメータを送信する．また，Q 学習エンジン開発アプリケーションは，Q 学習エンジンのシミュレータ機能を有し，シミュレーションによって Q-table を作成したり，適切な学習パラメータを決定したりできる．

6.3 内部状態のモニタリング

Q 学習エンジンの内部の様子は学習モード，ルールモード，マニュアルモードにおいて，Q 学習エンジンとパソコンのシリアル通信によりパソコン上の Q 学習エンジン開発アプリケーションでモニタリングすることができる．これにより，オンラインで Q 学習エンジンにおけるセンサ入力，状態，Q 値，選択された行動など，変化を観察して，Q 学習エンジンが正常に機能していることを確認できる．また，**図 6** のように Q-table の内容をパソコン上で点検してから，ほかの Q 学習エンジンに対して Q-table を送信できる．

7. Q 学習エンジンの応用

7.1 Q 学習エンジン単体での利用

家電製品に適切なセンサと Q 学習エンジンを組み込み，ユーザによる操作を教示入力として学習させることで，ユーザが好むように家電製品を制御する．

たとえば，ユビキタスコンピューティングの概念を提唱したワイザーは，サラ（登場人物の女性）の目覚めを感知し，コーヒーが沸くシナリオを紹介している[1]．ここで，家族のベッドに取り付けた圧力センサを入力とする Q 学習エンジンを，コーヒーメーカに組み込むことを考えると，コーヒーを飲む人が起床に合わせて作られたコーヒーの濃さ，温度，ブレンドなどを評価し，Q 学習エンジンにおいしかったかおいしくなかったかを何段階かで教えると，数日後には，Q 学習エンジンが家族全員の好みを習得し，それぞれの人が好むコーヒーを沸かすようになることが想定される．

そのほか，今回開発した Q 学習エンジンは衣服にも装着できるため，後述のように空調服のファン制御など，ウェアラブル機器への応用も可能である．

7.2 複数の Q 学習エンジンの連携

Q 学習エンジンはセンサ入力端子に，ほかの Q 学習エンジンの出力端子をつなぐことで，それらの出力

値を自らの状態の判定に利用し，複数の Q 学習エンジン間で連携を図ることができる．また，ほかのデバイスの入出力から状態を判定することも可能である．ここで，Q-table のサイズを制限するためには，通常の Q 学習と同様に，Q 学習エンジンの状態の定義において，どのデバイスの入出力に着目するかなど，ある程度の事前知識が必要となる．しかし，ユーザが好む Q 学習エンジンの連携は，ユーザが実際の環境下で状況を体験し，教示を与えることで達成されるのであり，設計者が事前に予測できるものではない．

たとえば，照明システムの知的制御に Q 学習エンジンを応用するならば，個々の照明器具に Q 学習エンジンを取り付け，隣接する Q 学習エンジンの出力をモニタリングさせる．これにより，全体の出力を抑えて省エネを図ったり，故障した照明器具の出力不足をカバーするような連携が期待できる．

また，小型で廉価な Q 学習エンジンを複数使用することで，簡易ホームネットワークを構築し，家電製品を連携させることができる．簡易ホームネットワークは，家屋の大規模な改修が不要であるため，老人の住宅ケアのためのリフォームに適している．さらに，Q 学習エンジンの学習機能によって，被介護者の症状に応じた家電製品の細かな制御が可能となる．

8. 空調服への応用例

8.1 市販の空調服と問題点

空調服とは衣服にファンを取り付け，ファンで外気を取り込んで体感温度を下げることにより，着衣者に快適な環境を作り出すというものである．人間には本来，発汗することにより身体を冷却するという機能が備わっている．この発汗による身体の冷却が有効に働くためには，汗の気化が不可欠である．空調服は衣服の中に身体の表面に対し，ほぼ平行に風を送ることにより，汗を効率良く気化させて身体の冷却を助けている．このため，クーラなどのように周りの環境に影響を与えることがなく，個人ごとに最適な服内環境を作り出すことができ，省エネの効果もある．ちなみに，現在市販されている空調服の風量は 2 段階に調節可能で，単三型乾電池 4 本で約 5 時間使用できる[18]．

8.2 Q 学習エンジンの構成

人間の身体からの発熱量は，周囲の環境や運動などにより，たえず変化している．市販の空調服では，そのような発熱量の変化に対し，ユーザが手動でファンの速度を調整しなければならない．そこで，本論文では，ユーザの負担を軽減するため，空調服のファン速度の制御に Q 学習エンジンを適用した．

表 4 状態空間の定義
Table 4 Definition of states.

状態	体感温度の範囲	状態	体感温度の範囲
s_0	$0 \leq T < 12$	s_8	$26 \leq T < 28$
s_1	$12 \leq T < 14$	s_9	$28 \leq T < 30$
s_2	$14 \leq T < 16$	s_{10}	$30 \leq T < 32$
s_3	$16 \leq T < 18$	s_{11}	$32 \leq T < 34$
s_4	$18 \leq T < 20$	s_{12}	$34 \leq T < 36$
s_5	$20 \leq T < 22$	s_{13}	$36 \leq T < 38$
s_6	$22 \leq T < 24$	s_{14}	$38 \leq T < 40$
s_7	$24 \leq T < 26$	s_{15}	$40 \leq T$

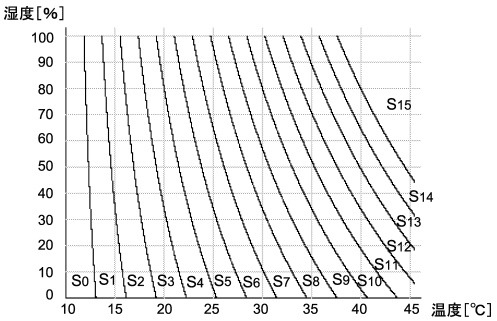

図 7 状態空間図
Fig. 7 Classification of state spaces.

センサ入力と状態

ユビキタス Q 学習における状態 s_i は，温度センサから得られた温度 t［°C］と，湿度センサから得られた湿度 h［%］をもとに式 (5) のミスナールの体感温度 T［°C］を算出し，**表 4** に示す 16 段階で定義した．また，温度と湿度に対する状態空間を**図 7** に示す．

$$T = t - \frac{1}{2.3}(t - 10)\left(0.8 - \frac{h}{100}\right) \tag{5}$$

ユーザ教示入力とファン制御

ファンの回転速度は，Q 学習エンジンからの出力パルス幅により，16 段階で制御される．一方，4.2 節の Q 学習における出力 a_j は速度を「上げる」,「下げる」,「変えない」の 3 種類とする．このため，ユーザ教示入力は，ファンの速度を上げるか，下げるかの 2 種類のみでよく，Q 学習エンジンのユーザ教示用入力端子に接続された 2 個のプッシュボタンで実行する．

8.3 実験による評価

計算機シミュレーション

Q 学習エンジンを空調服のファン制御に応用した場合について，通常の Q 学習とユーザ教示のある Q 学習の効果を計算機シミュレーションで比較した．

計算機シミュレーションでは，ランダムな温度・湿度から，目標とする状態（ゴール）に到達した回数を計測した．ゴールに到着するとランダムな温度・湿度

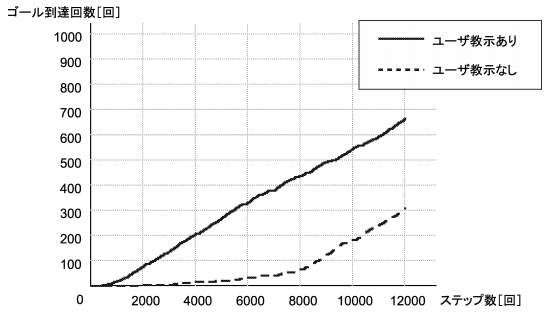

図 8 計算機シミュレーション結果
Fig. 8 Results of computer simulation.

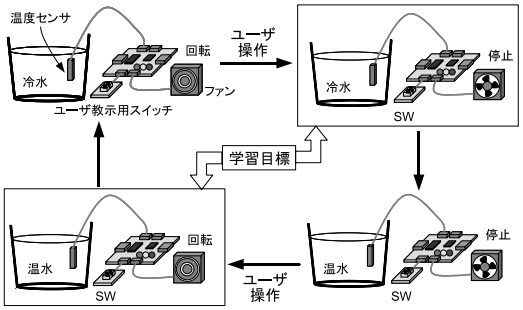

図 9 Q 学習エンジンの官能検査
Fig. 9 Sensory inspection of Q-learning engine.

の状態に移行し，再びゴールにたどり着くように Q 学習を繰り返した．ここで，Q 学習で 1 つの Q 値を更新することを 1 ステップとし，ユーザ教示ありの場合は，100 ステップに 1 度の割合でユーザ教示を与えた．Q 学習のパラメータは，事前の実験から，ともに学習率 $\alpha = 0.4$，割引率 $\gamma = 0.9$ とし，方策は $\epsilon = 0.2$ の $\epsilon - greedy$ 手法[7] を用いた．また，身体からの発熱量と発汗量は一定とし，服内の湿度は温度と飽和水蒸気量から決まり，温度の変化と排気量，および，排出量はファンの回転速度に依存するものとした．

Q 学習を計 12,000 ステップ実行した結果を**図 8** に示す．図 8 より，ユーザ教示のある Q 学習は，教示のない場合に比べて，ゴール到達回数が非常に多いことを確認できる．

官能検査

被験者 8 名（男子の大学生）に Q 学習エンジンの操作方法を説明した後，**図 9** に示すように，温度センサが冷水に浸かっているときはファンが停止し，温水に浸けたときにはファンが回転するように教示を与えてもらう．温水と冷水での教示を 6 回ずつ繰り返した後に，各被験者が感じた学習の効果を 0〜10 点で評価してもらった．ただし，8 名の被験者を 2 グループに分け，被験者たちには知らせずに，一方のグループは Q 学習エンジンを学習モードに，他方のグループでは学習機能のないマニュアルモードに設定しておいた．

被験者 8 名（A〜H）の採点結果を**表 5** に示す．表 5 によれば，学習モードで使用したグループによる評価数が高いといえる．さらに，個人差の影響を誤差因子，学習の有無を変動因子とし，表 5 の結果を分散分析で解析した結果を**表 6** に示す．F 値から，学習の有無が評価点に影響を与えている確率は 99［%］以上と推定され，個人差は無視できるといえる[19]．

表 5 官能検査結果
Table 5 Result of sensory inspection.

	学習ありグループ				学習なしグループ			
被験者	A	B	C	D	E	F	G	H
採点（0〜10 点）	8	6	8	7	1	2	3	2

表 6 分散分析結果
Table 6 Result of analysis of variance.

変動要因	平方和	自由度	平均平方	F 値
学習有無	13.78	1	13.78	17.41
誤差因子	4.75	6	0.79	

8.4 考　　察

一般に学習アルゴリズムの有用性の検証は難しく，特にユビキタス環境という将来の環境における評価を完全に行うことは難しい．そのようななか，今回 Q 学習エンジンの有用性について計算機シミュレーションおよび複数の被験者による官能検査と分散分析によって，ある程度の実証が行えたものと考える．また Q 学習エンジンの小ささや，LED による Q 学習エンジンの動作の表示，マニュアルモードの存在など，実用的な意味でのユビキタス学習の要件も，おおむね達成できたと考えられる．

今後，空調服のファンのモータ制御で得られた知見は，ユーザの好みを学習する様々な家電製品の開発のほか，Q 学習エンジンによるウェアラブル・ロボットのアクチュエータの制御[20] にも活用できることが考えられ，今後の実環境での評価が期待される．

9. おわりに

本論文では，強化学習の手法の 1 つである Q 学習をベースにユーザ教示とデータ通信により Q-table の更新が行える Q 学習エンジンをマイクロコントローラを用いて実装し，そのハードウェア構成や学習アルゴリズムについて述べた．また，空調服のファンの制御に応用し，その有効性を実験によって確認した．

今後の課題として，複数の Q 学習エンジンによる連携の応用例において有用性を確認することのほか，適用例に応じたユーザによる教示方法の工夫があげられる．たとえば，ハンズフリーな教示方法としては，

カメラと加速度センサを用いたポインティング[21] がある．また，ユーザの意識をともなわない教示方法としては，脳波や筋電の利用などが考えられる．さらに，無線による通信や入出力端子の多様化など，Q 学習エンジンの機能のさらなる強化も必要である．

謝辞 本研究の一部は，(財）人工知能研究振興財団，および，文部科学省特定領域研究「情報爆発のための装着型入出力デバイスを用いた情報操作方式」（19024056）の助成を受けた．ここに記して深謝する．

参 考 文 献

1) Weiser, M.: The Computer for the Twenty-first Century, *Scientific American*, Vol.265, No.3, pp.94–104 (1991).
2) Wellner, P., Machay, E., Gold, R., Weiser, M., et al.: Computer-Augmented Environments Back To The Real World, *Comm. ACM*, Vol.36, No.7, pp.271–278 (1999).
3) http://www.xbox.com/Products/wireless/SensorNetwork.htm
4) http://www.smart-its.org/
5) 早川敬介，塚本昌彦，寺田 努，義久智樹，岸野泰恵，柏谷 篤，坂根 裕，西尾章治郎：ユビキタスコンピューティングのためのルールに基づく入出力制御デバイス，ヒューマンインターフェース学会論文誌，Vol.5, No.3, pp.341–354 (2003).
6) 岡田量太，田川聖治，塚本昌彦：ユビキタス Q 学習エンジンの設計と実装，第 50 回システム制御情報学会研究発表講演会 (2006).
7) Sutton, R.S. and Barto, A.G.: *Reinforcement Learning*, The MIT Press (1988).
8) http://www.robocup.or.jp/
9) http://www.u-mart.org/
10) 中村恭之：実機ロボットリーグの現状と今後の課題，日本ロボット学会誌，Vol.20, No.1, pp.11–14 (2002).
11) 酒井 勝：ロボットの直接教示，日本ロボット学会誌，Vol.13, No.5, pp.627–628 (1995).
12) 片上大輔，山田誠二：対話的進化ロボティクスの観測に基づく教示の設計，システム制御情報学会論文誌，Vol.16, No.6, pp.279–286 (2003).
13) 丹 康雄（監修），宅内情報家電・放送高度化フォーラム（編）：ホームネットワークと情報家電，ユビキタス情報シリーズ，オーム社 (2004).
14) 富田浩司，三木光範，廣安知之：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，第 13 回自律分散システム・シンポジウム，pp.27–32 (2001).
15) 前田英作，南 泰浩，堂坂浩二：妖精・妖怪の復権—新しい「環境知能」像の提案，「50 年後の情報科学技術をめざして」記念論文，情報処理，Vol.47, No.6, pp.624–640 (2006).
16) 村上陽一郎：安全と安心の科学，集英社 (2005).
17) 浅田 稔，野田彰一，俵積田健，細田 耕：視覚に基づく強化学習によるロボットの行動獲得，日本ロボット学会誌，Vol.13, No.1, pp.68–74 (1995).
18) http://www.9229.co.jp/
19) 菅 民郎：Excel で学ぶ実験計画法，オーム社 (2002).
20) 川村貞夫：ウェアブル・ロボットのアクチュエータとシステム，日本ロボット学会誌，Vol.20, No.8, pp.783–786 (2002).
21) 小川樹幸，塚本昌彦，義久智樹，西尾章治郎：カメラと加速度センサを用いたポインティング方式の設計と実装，ウェアブルコンピューティング研究会研究報告，Vol.1, No.1, pp.78–85 (2005).

(平成 18 年 10 月 31 日受付)
(平成 19 年 4 月 6 日採録)

岡田 量太

2005 年神戸大学工学部電気電子工学科卒業．2007 年同大学院自然科学研究科電気電子工学専攻修士課程修了．同年（株）日立製作所（Hitachi）に入社．現在，(株）カシオ日立モバイルコミュニケーションズに出向中．ユビキタスコンピューティングに興味を持つ．

田川 聖治（正会員）

1991 年神戸大学大学院工学研究科修了．1993 年神戸大学工学部助手．2005 年同学部助教授．2007 年近畿大学理工学部教授となり，現在に至る．メタ戦略・進化型計算とその応用，ユビキタス機械学習に関する研究に従事．博士（工学）．計測自動制御学会，システム制御情報学会，電気学会，IEEE 各会員．

塚本 昌彦（正会員）

1987 年京都大学工学部数理工学科卒業．1989 年同大学院工学研究科修士課程修了．同年シャープ（株）入社．1995 年大阪大学工学部情報システム工学科講師，1996 年同助教授を経て，2004 年神戸大学工学部電気電子工学科教授となり，現在に至る．工学博士．ウェアラブルコンピューティングとユビキタスコンピューティングの研究に従事．ACM，IEEE 等 8 学会の会員．