居住者の嗜好を少ないインタラクションで推定するスマートホームシステム

辰巳 公太[1]　エルデーイ ヴィクトル[1]　水本 旭洋[1]　山口 弘純[1]　東野 輝夫[1]

概要：空調や照明といった環境制御において，スマートホームシステムが居住者の嗜好を理解するためには，制御を行った時刻や位置，感情や気分，同居者の存在など様々な周辺コンテキストに影響される居住者の嗜好をシステム側が獲得・理解し，サービス提供のタイミングや制御方法に反映させることが望ましい．そのためには，制御とそれに対する居住者の反応（フィードバック）を取得し，強化学習等でシステムを居住者に適応させることが考えられる．しかし，居住者のフィードバックを能動的に取得する場合に生じるシステムとのインタラクションはなるべく効率的かつ居住者に負担のない形で行う必要がある．環境制御を学習すると同時に，インタラクションのベストタイミングや内容も最適化できるシステムの実現が望まれる．本研究では，環境制御に対する居住者の嗜好を強化学習により理解するシステムにおいて，嗜好の獲得のためにシステムが行なう能動的なインタラクションに対する居住者の嗜好も同時に学習する手法を提案する．提案手法では，環境制御に対する嗜好をスマートスピーカーやスマートフォンのようなインタフェースを介した質問に対するフィードバックで獲得し，それを報酬とした強化学習を行う．その際，フィードバック獲得のためのインタラクションのタイミングや内容に対するフィードバックも同時に獲得し，次の制御において質問をするか否かの振舞いを決定する．この目的のため，Q 学習に基づくインタラクション制御機能を有したシステムアーキテクチャを設計している．同システムを実装したシミュレータを用いた簡易実験を行った結果，学習達成度 80%以上達成した時に，大幅なインタラクション回数の削減に成功した．

1. はじめに

IoT 関連技術の発展や，スマートスピーカーやヘルスモニタリングデバイスなどのコモディティ化と普及により，多くの企業がスマートホーム市場に参入し，スマートホームに関する技術や製品のさらなる高度化と普及が期待されている．中でも，室温や照度といった環境センシング情報と，居住者の位置や行動，気分などのセンシング情報を組み合わせ，居住者の状況や好みに応じて家電や空調・照明や音楽などを制御し，エネルギーを抑制しながら適度に自動化された快適な生活を実現するスマートホームシステム技術の実現が期待されている [1].

そういったシステムの実現に向け，家庭内コンテキストを取得し環境制御に活用する手法が多く提案されてきている．文献 [2] では，宅内センシングで居住者の行動や温度・湿度・照度などのコンテキスト情報を認識し，事前に設定したルールに基づいてサービスを提供するようなシステムを提案している．また，ルールベースの手法において居住者に応じたルールを適切に設定する手間や煩雑さを避けるため，学習ベースの手法も多く提案されている．文献 [3] では，スマートホームにおいて，コンテキストの組み合わせで行動パターンを表現しオンライン学習するシステムである LaPlace を提案している．

環境制御に対する居住者の好み（嗜好）は，制御を行った時刻や位置，感情や気分，同居者の存在など様々な周辺コンテキストに依存する．したがって，システムが環境に対する居住者の好みを理解するためには，それらのコンテキストをシステム側が獲得・理解し，サービス提供のタイミングや制御方法に反映させることが望まれる．そういった目的に対し，制御とそれに対する居住者の反応（フィードバック）を取得し，強化学習などでシステムを徐々に居住者に適応させることが考えられる．しかし，それらのインタラクション自体も適切なタイミングや内容で行われることが望ましい．例えば，時刻や環境に応じた室温や照度設定の好みを学習する際には，なるべく効率的かつ居住者に負担をかけないよう，居住者に対するインタラクション（制御に対する満足度の質問）を最適化できるシステムで

[1] 大阪大学 大学院情報科学研究科
Graduate School of Information Science & Technology, Osaka University,
1-5 Yamadaoka, Suita, Osaka 565–0871, Japan

あることが望ましい．

本研究では，環境制御に対する居住者の嗜好を強化学習により理解するシステムにおいて，嗜好の獲得のためにシステムが行なう能動的なインタラクションに対する居住者の嗜好も同時に学習する手法を提案する．提案手法では，環境制御に対する嗜好をスマートスピーカーやスマートフォンのようなインタフェースを介した質問に対するフィードバックで獲得し，それを報酬とした深層強化学習を行う．その際，フィードバック獲得のためのインタラクションのタイミングや内容に対するフィードバックも同時に獲得し，次の制御において質問をするか否かの振舞いを決定する．この目的のため，Q 学習に基づくインタラクション制御機能を有したシステムアーキテクチャを設計している．同システムを実装したシミュレータを用いた簡易実験を行った結果，学習達成度 80%以上達成した時に，大幅なインタラクション回数の削減に成功した．

2. 関連研究

センシングにより得られた宅内のデータに対し，位置推定や行動認識，感情推定などを行い，見守りや家電制御，エネルギー削減といったサービスに活用する事例がこれまでに多く提案されている．

文献 [2] では，宅内における環境センシングデータから得られるデータなどから家庭内のコンテキストを定義し，あらかじめ設定したルールに基づいてサービスを提供するルールベースのコンテキスト把握技術を提案している．文献 [4] では，行動や健康に関する変化を検出するパターン認識モデルを提案しており，異常の早期検知を目指している．隠れマルコフモデルや行動履歴からの日常生活の異常検知とバイタルサインの将来予測を行っており，ファジールールベースのモデルから最終予測を出力している．また，文献 [3] では，コンテキストの組み合わせで行動パターンを表現し，行動モニタリングや環境変化検知機能を有するスマートホームにおけるオンライン学習システムを提案している．

また，能動的に人間行動を理解するシステムの構築に向けた研究も行われつつある．文献 [5] では，スマートホームにおいて人間からの入力に応じて動作する「召使」ではなく，「執事」としての役割をする人間中心型のスマートホームエネルギー管理システムを提案している．このシステムは人間行動の認知と電力使用パターンの相関からユーザの電力利用デマンドを発見し最適スケジューリングを支援する方法を提案している．文献 [6] では，スマートホームにおけるユーザインタフェースと設定の複雑化に伴いそれらの学習に費やす時間が増加しつつあることに着目し，ユーザが次に何をしたいかを予測するモデルを協調フィルターとコンテンツベースの推薦システムのハイブリッドモデルで構築している．文献 [7] では，自動化や快適性を実現す

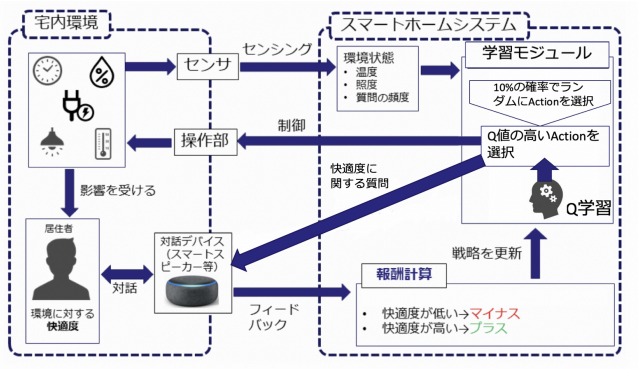

図 1　システム概要

るホームオートメーションシステムにおいて，十分な情報を有さない音声による意思伝達に対し，センサデータから推定される状況から意思決定を行うシステムフレームワークを設計開発しており，マルコフロジックネットワーク用いた知識表現と推論機構を実現している．スマートホームではないものの，文献 [8] では，ビルディングの HVAC システムを対象とし，エネルギー消費を抑制する手法を提案している．従来のルールベースの制御ポリシでは温熱環境の流動性に対応することが難しいとし，多数の被験者から収集したデータに基づき，複数のコンテキストを学習するモデルを組み入れた深層強化学習を用いた学習型の制御ポリシを提案している．この手法では，システムが推定した人の快適な温度を維持しながら消費電力の最小化を目指しており，学習におけるフィードバック回数を削減している．

これらに対し，提案手法では，システムがユーザの嗜好を学習する際に発生するインタラクションに対する受容のタイミングやインタラクションの内容もユーザ毎に異なるため，そのユーザに適したインタラクション最適化機構を備えることが望ましいといった発想に基いている．これにしたがい，強化学習によるオンライン学習型の環境制御フレームワークにおいて，ユーザ（居住者）とシステムのインタラクションの効率性向上に着目し，その嗜好に応じたインタラクションの最適化機構をフレームワークに組み入れた新しいアーキテクチャを提案している点で前述のアプローチとは異なる．

3. 提案する制御システムの概要

図 2 に提案システムのアーキクチャを示す．

本研究では簡単のため．室温および照度の制御を例題として用いるが，手法はこれらのコンテキストに依存しない．提案システムでは，対象空間において環境（室温および照度）をセンシングにより取得するのためのセンサ，制御対象となるアクチュエータ（ここではエアコンや照明器具），および環境制御に対し居住者にその快適度を質問し，回答を得るための音声デバイスなどの対話デバイスが宅内にあるとする．また，スマートホームシステムは対話デバイス

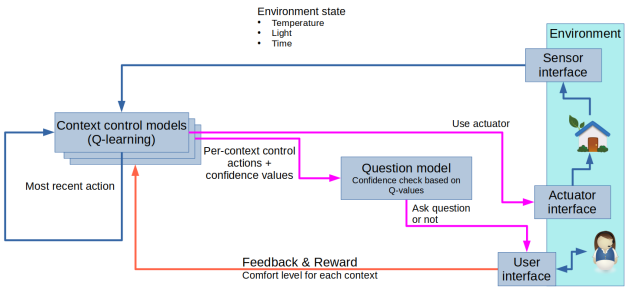

図 2　モデルの構築フロー

を介してユーザに質問を行うことができ，質問に対し，制御された結果の室温と照度の快適度，ならびに質問されたことに対する許容度（質問の快適度とよぶ）を回答する（これをフィードバックとよぶ）．室温，照度および質問のいずれにおいても快適度は [0,1] の実数であるとし，なんらかの形で居住者の回答がこれらの数値に変換されるものとする．本論文の実験では簡単のため，0（不快）あるいは 1（快適）の 2 値のフィードバックが得られるものとする．

なお本研究で想定する居住者は，質問されることに対し，システムが学習するまでにある程度は許容できるものの，なるべく質問回数が少ないことを望むような居住者であるとする．本来であれば，居住者の許容度は，質問されるタイミング（多忙な時刻や毎日同じような状況で質問されることは望まない，など）や質問の時間間隔，居住者の状態（邪魔されたくない状態）などに依存し，提案手法ではそれらのコンテキストも考慮した学習も可能であるが，本論文では質問回数のみに着目する．

フィードバックから得られる温度，照度の快適度に対し，Q 学習における State（室温および照度）および Action（それらの制御）に対する報酬計算を行う．詳細は次章で述べるが，室温および照度それぞれに対し，State と Action に対する Q 値を Q テーブルに保存し，快適度が一定以上であれば報酬として 1 を与え，Q 値を更新する方法を採用する．最後に，現在の State に対し，温度または照度の Q 値がどのような Action においても向上していない場合が起きないために，10%の確率でランダムに Action を行いフィードバックを得て Q 値の更新を行う．

4. 提案手法

提案手法では，スマートホームのコンテキストに応じて，各環境（温度や照度など）に対する制御を出力する環境制御モデルにより，居住者の嗜好の学習，および，嗜好にあった環境制御を実現する．図 2 は，環境制御モデルの構築フローを表している．環境制御モデルでは，3 節で述べたように，Q 学習を利用しており，温度や照度などの各環境を State，各 State に対して出力される制御を Action，居住者へ質問を行うと判定された Action に対する快適度をフィードバックとして居住者から取得し，環境制御モデルが保持する Q テーブルの Q 値を更新していくことで，居住者の嗜好を学習する．

4.1 State と Action

本稿では，温度，照度，時間の 3 要素を State s として以下のように定義する．

$$s = (Temperature, Light, Time) \quad (1)$$

State s に対して，環境制御モデルにより出力される制御を Action a とする．本手法では，$Temperature$ や $Light$ などの環境情報ごとに制御モデルを構築することを想定しており，温度に関する環境制御モデルの場合，s の温度 $Temperature$ と時間 $Time$ を用いて，照度に関する環境制御モデルの場合，s の照度 $Light$ と時間 $Time$ を用いて，Action a がそれぞれのモデルから出力される．出力される Action は，例えば，温度に関する環境制御モデルの場合，「温度を x 度下げる」というようなものである．

4.2 環境制御モデル

本研究では，Q 学習に基づいて環境制御モデルを構築する．Q 学習は，State s を入力として，出力される Action a に対する，フィードバックから報酬 r を計算し，State と Action に対する状態行動価値 **Q 値**を更新していく手法である．State と Action に対する Q 値は，モデルにおいて，Q テーブルと呼ばれるテーブルで管理される．

本環境制御モデルでは，任意の State において，実行可能な Action の集合（以下，アクションスペース）の中から，Q 値が最も高い Action を出力し，スマートホームのアクチュエータを通じて環境を制御する．また，学習が十分に行われていない Action に関して，フィードバックを得るために，環境制御モデルは，10%の確率でランダムな Action を生成するようにしている．出力した制御に対する居住者のフィードバックが好ましくない場合には，アクチュエータに制御の停止命令が送信される．

4.2.1 Q 値の更新方法

学習開始時には，Q テーブルの全ての Q 値は 0 に設定されており，Q 値はフィードバックにより，以下の方法で更新される．

$$Q(s,a) := Q(s,a) + \alpha(r + \gamma \max_{a' \in A}(Q(s',a')) - Q(s,a)) \quad (2)$$

ここで，α は学習率 (learning rate)，r は報酬，γ は割引率 (discount factor)，A はアクションスペースを表す．5 章の評価実験では，居住者モデルは確定的であるため，学習率を 1 に設定している．割引率は，0 に近づくほど，短期報酬を優先するが，1 に近く設定するほど，長期報酬を目指すようになる．本システムでは，一つの Action で報酬が

与えられるとは限らないため，0.5 に設定する．報酬 r は，実数で表される居住者のフィードバックが一定以上の場合に 1 が与えられる．本稿の実験においては，簡単化のため，0（不快）か 1（快）の 2 値のフィードバックが得られるものとして報酬 r を計算している．

4.3 質問判定手法

4.2 節で述べたように，居住者の嗜好を学習するためには，Q 値の更新が必要であるが，これには居住者からのフィードバックが必要となる．しかしながら，質問回数が多くなると，居住者の負担が大きくなるため，質問回数や頻度を減らす仕組みが必要だと考えられる．4.2 節で述べた通り，環境制御モデルは，State s，および，Action a に対する，Q 値を Q テーブルで管理している．この値は報酬として居住者の快適度が最大となるように学習されており，居住者の快適度を示す 1 つの指標と見做すことができる．そこで，本提案手法では，居住者への質問を行うか否かを，Q 値に基づき判定することで，質問回数や頻度の抑制を実現する．

質問判定では，式 3 で示すように，任意の State に対して，モデルにより出力された Action の Q 値と，その State に対するアクションスペースの平均 Q 値を比較して平均 Q 値が下回る場合か，平均 Q 値が 0 である場合に，居住者の嗜好を十分に学習できていないと判定し，居住者に質問を行うことで，フィードバックを取得する．それ以外の場合には，質問が行われないため，Q 値は更新されない．

$$Q(s,a) > m * average_{a' \in A}(Q(s,a')) \ || \\ average_{a' \in A}(Q(s,a')) = 0 \quad (3)$$

ここで，m は，出力された Action の Q 値が，平均を「著しく」超えるかを判定するための乗数である．様々な値を試した結果，最適値として 1.01 を設定した．なお，提案手法では，環境ごとに環境制御モデルを構築しているため，上記の判定は，モデルごとに行われる．

5. 評価実験

提案手法では，居住者の嗜好推定を行うために，複数の環境制御モデルを構築するとともに，質問判定手法により，フィードバックを得るためのインタラクション回数を削減している．本実験では，シミュレーションを用いて，提案手法による質問回数（インタラクション数）の削減効果と，学習モデルの精度について，毎回質問する場合，および，50 %の確率で質問を行う場合と比較実験を行った．

5.1 実験環境

本シュミレーションでは，State として，温度，照度，時間を収集でき，また，Action によって，温度，照度を制御可能なスマートホーム環境を想定する．State が取りうる範囲は，温度は 10°C から 40°C の範囲，照度は 11 段階の値で 0 から 10 の範囲，時間は 10 分刻みで表される．Action として，温度の環境制御モデルからは，-10°C から+10°C までの範囲で温度を制御する Action が，照度の環境制御モデルからは，-5 から+5 までの範囲で照度を制御する Action が出力される．State の取りうる範囲を超えるような Action が出力された場合には，範囲の下限値または上限値に合わせた Action に変更される．

表 1 温度の快適度設定

温度 °C	時間	報酬
16°C〜18°C	0:00〜8:00	+1
20°C〜22°C	8:00〜20:00	+1
16°C〜18°C	20:00〜24:00	+1

表 2 照度の快適度設定

照度	時間	報酬
0〜3	0:00〜6:00	+1
3〜6	6:00〜18:00	+1
6〜10	18:00〜23:00	+1
0〜2	23:00〜24:00	+1

5.2 居住者の快適度モデル設定

本実験では，居住者のフィードバックをシミュレートする快適度モデルを実装した．快適度モデルは，システムから受け取った State と Action の組み合わせに対して，フィードバックを環境制御モデルに渡す．ここでは簡単化のため，フィードバックはそのまま報酬として扱われるように，居住者の嗜好を満たしている場合には 1，それ以外は 0 が出力される．

温度の快適度計算は表 1 を参照して行う．例えば State と Action の組み合わせの結果として，次の State が，時間 14:00，温度 21°C の場合，報酬が 1 加算される．

同様に，照度の快適度計算は表 2 を参照して行う．例えば State と Action の組み合わせの結果として，次の State が，時間 10:00，照度が 11 段階のうち 4 の値となれば報酬が 1 加算される．

5.3 実験手法

本実験では，学習フェーズとテストフェーズという 2 つのフェーズにより，学習モデルの精度およびインタラクション回数の削減効果を評価する．

本実験では，任意の State において環境制御モデルから出力された Action に対して，居住者からフィードバックを受け取り，Q 値を更新するまでの一工程を 1 step とする．また，100 step を 1 episode として，学習およびテストを行う．以下では，それぞれのフェーズにおける実験手

順について詳細を述べる．

5.3.1 学習フェーズ

学習フェーズでは，1000 回 episode を繰り返すことで，環境制御モデルの学習を行う．各 episode において初期 State はランダムに生成される．そして，生成された初期 State を各環境制御モデルに入力し，State に対して最も Q 値が高くなる Action を出力される．学習フェーズにおいては，Q 値の計算が十分ではない Action についても学習を行いやすくするために，10%の確率でランダムに Action を出力する．そして，出力された Action を回答判定モデルを用いて居住者に質問するか判定する．質問すると判定された Action については，節で定義したモデルにより，フィードバック（報酬）が決定され，対応する環境制御モデルの Q 値が更新されることで，1 step が終了する．次の step において，環境制御モデルに入力される State は，前の step において，出力された Action の実行によって遷移する先の State とした．例えば，前の step で，State が 30°C，Action -5°C が出力されていた場合，次の step では，State 25°C が環境制御モデルに入力される．

上記の処理を 100 step 繰り返し，1 episode が完了した際に，その episode における，合計報酬（Total Reward）と平均報酬（Avg. Reward）の計算を行う．

5.3.2 テストフェーズ

テストフェーズでは，学習フェーズと同様に，500 回 episode を実行し，それぞれの episode において，学習済のモデルに対する合計報酬と平均報酬を計算する．このフェーズにおいては，各環境制御モデルは学習済みであると考え，ランダムな Action 出力は行わず，最も Q 値が高くなる Action のみが出力される．

5.4 評価指標

質問判定モデルの評価を行うためのベースラインとして，毎回質問を行うモデルと，50%の確率で質問を行うモデルとの評価実験を 5.3 節で説明した手法で行う．これらを評価する指標として，学習モデルの精度を用いる．学習モデルの精度は 1 episode において，各 step で決定された報酬の平均である AvgReward を利用する．AvgReward は快適度モデルのフィードバックから計算されるため，その割合は学習の達成度を表す．これを本シミュレーションにおける評価指標として用いる．

5.5 モデルが学習した快適度の評価

本実験は 5.3.1 節で述べたモデルに基づき嗜好推定を行う．実験にて学習を行ったモデルの Q 値に従い，出力した Action の精度評価を図 3 に示す．縦軸は State における時間を表し，上図の横軸が温度，下図の横軸が照度を表す．この図 3 は縦軸と横軸それぞれを組み合わせた State における最適な Action の結果が，快適度モデルとどの程度差

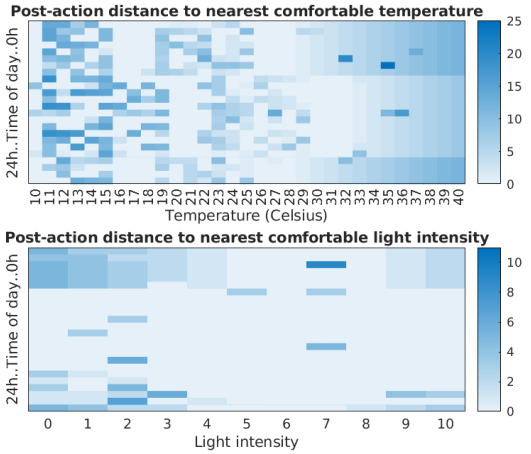

図 3 学習したモデルの評価

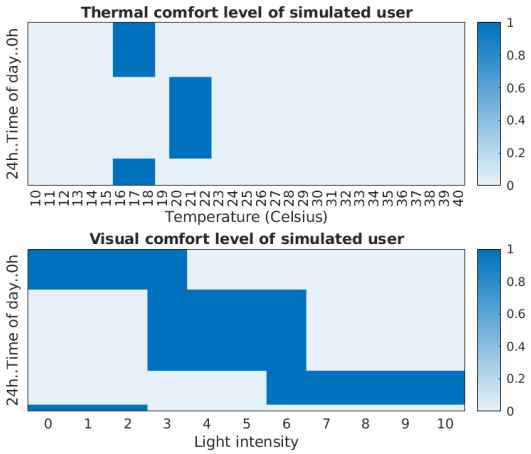

図 4 想定した快適度モデル

が開いているかを値の度合いとして表している．値が大きいほど学習したモデルの Q 値に従い出力した Action が快適度モデルから離れており，Action の精度が悪いことを意味する．逆に値が小さいほど快適度モデルと Action の差が少なく，精度が良いことを意味する．図 4 は 5.3.1 節で述べたモデルの嗜好を表す．図 3 と図 4 を比較すると，図 4 の快適度に近づく Action の精度が上がっているため，学習モデルの精度は居住者の快適度を反映することがわかる．そのため学習の達成度を表す AvgReward は居住者の快適度を意味する．

5.6 実験結果

毎回質問を行うモデルの学習結果を図 7, 50%の確率で質問を行うモデルの学習結果を図 6, 質問判定有りモデルの学習結果を図 5 にそれぞれ示す．

episode 数 300 回程度で，評価指標である各要素の AvgReward が 80%以上を達成した．また図 7 における NumQuestion の変動より質問回数が大幅に削減されていることが確認できた．

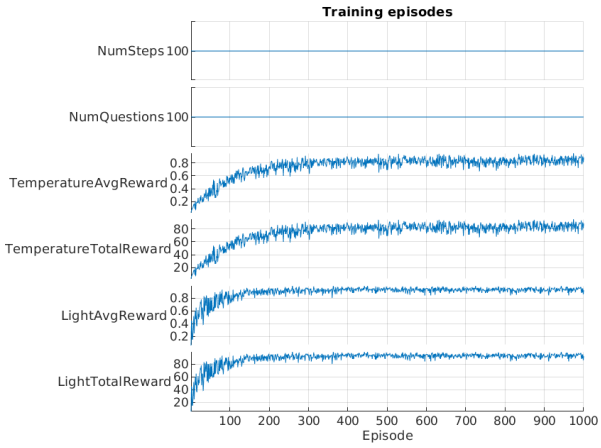

図 5　毎回質問を行うモデルの学習結果

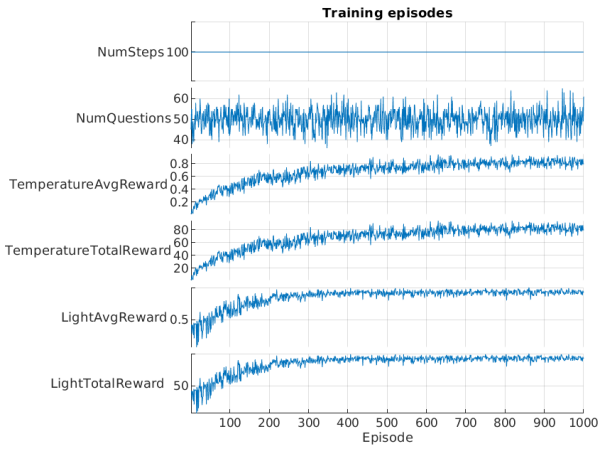

図 6　50%の確率で質問を行うモデルの学習結果

50%の確率で質問を行う学習モデルは，episode 数 500 回の程度で AvgReward が 80%以上を達成し，質問判定有りのモデルと比べ必要学習数が増加したことが判明した．よって質問の内容を考慮せずに質問を行うモデルより，質問判定有りモデルの方が少ない学習回数で居住者の嗜好推定を行えたと言える．

一方，毎回質問を行うモデルは質問判定モデルありと同じ episode 数で AvgReward80%以上を達成した．このことから質問判定ありのモデルは質問内容の選択により，学習精度を向上させたことが確認できた．

またそれぞれの学習モデルのテスト結果を図 10, 図 8 図 9 にそれぞれ示す．これらの結果より，3 手法とも AvgReward が 95%以上の精度となっており，十分に居住者の快適度を学習できたと言える．

6.　おわりに

本稿では，住空間における居住者の嗜好を少ないインタラクションで推定するための手法を提案した．提案手法では，Q 学習を使用した居住者の嗜好推定手法に質問判定モデルを追加することにより，居住者からのフィードバック取得について，単純に毎回質問を行う場合や，不定期に質問を行う場合と比べて，質問回数を削減することができる．評価実験において，提案手法により，学習モデルの精度が 80%以上の時，学習を行う際に必要な質問の判定を行えて

図 7　質問判定有りモデルの学習結果

図 8　毎回質問を行うモデルのテスト結果

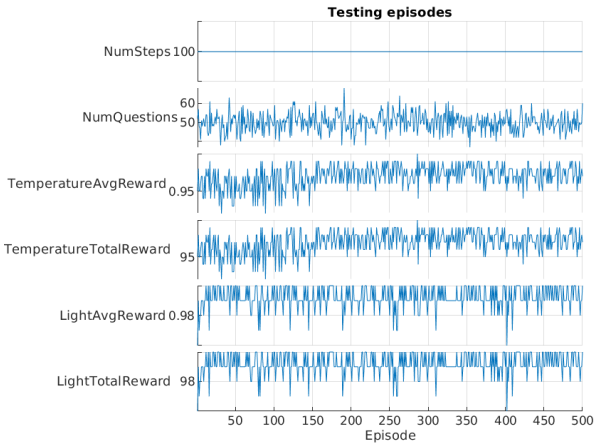

図 9　50%の確率で質問を行うモデルのテスト結果

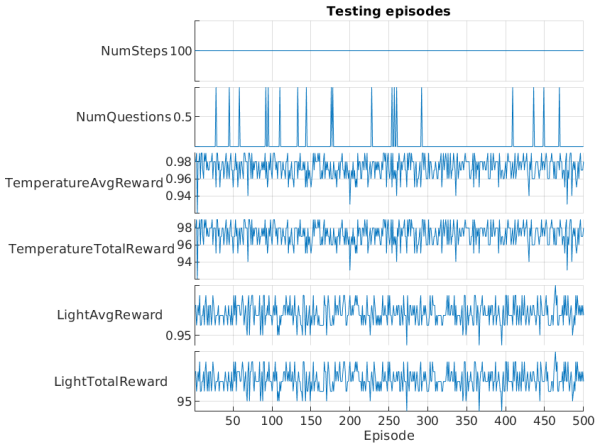

図 10 質問判定有りモデルのテスト結果

いることを示した．

今後の課題として，温度や照度だけではなく，様々なコンテキストを組み合わせた学習が行えるように，深層学習を用いた学習モデルを構築する予定である．また，提案手法を実環境に実装し，より現実的な評価を行う予定である．

参考文献

[1] Krishnan, N. C. and Cook, D. J.: Activity recognition on streaming sensor data, *Pervasive and Mobile Computing*, Vol. 10, pp. 138 – 154 (2014).

[2] Meng, Z. and Lu, J.: A Rule-based Service Customization Strategy for Smart Home Context-Aware Automation, *IEEE Transactions on Mobile Computing*, Vol. 15, No. 3, pp. 558–571 (2016).

[3] Lago, P., Roncancio, C. and Jimenez-Guarin, C.: Learning and managing context enriched behavior patterns in smart homes, *Future Generation Computer Systems*, Vol. 91, pp. 191 – 205 (2019).

[4] Forkan, A. R. M., Khalil, I., Tari, Z., Foufou, S. and Bouras, A.: A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living, *Pattern Recognition*, Vol. 48, No. 3, pp. 628 – 641 (2015).

[5] Chen, S., Liu, T., Gao, F., Ji, J., Xu, Z., Qian, B., Wu, H. and Guan, X.: Butler, Not Servant: A Human-Centric Smart Home Energy Management System, *IEEE Communications Magazine*, Vol. 55, No. 2, pp. 27–33 (2017).

[6] Chen, H., Xie, X., Shu, W. and Xiong, N.: An Efficient Recommendation Filter Model on Smart Home Big Data Analytics for Enhanced Living Environments, *Sensors*, Vol. 16, No. 10, p. 1706 (2016).

[7] Chahuara, P., Portet, F. and Vacher, M.: Context-aware decision making under uncertainty for voice-based control of smart home, *Expert Systems with Applications*, Vol. 75, pp. 63 – 79 (2017).

[8] Wei, T., Yanzhi Wang and Zhu, Q.: Deep reinforcement learning for building HVAC control, *2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1–6 (2017).