ユーザとの対話に基づく段階的な行動決定モデルの獲得

稲 邑 哲 也^{*1*2} 稲 葉 雅 幸^{*2} 井 上 博 允^{*2}

Incremental Acquisition of Behavior Decision Model based on Interaction between Human and Robots

Tetsunari Inamura^{*1*2}, Masayuki Inaba^{*2} and Hirochika Inoue^{*2}

In this paper, we propose a novel method for personal robots to acquire autonomous behaviors based on interaction between users and robots. This method have two advantages, first is loadless for users by comparison with users' teaching, second is that robots can acquire behaviors as users wish by comparison with autonomous learning. In this method, robots store sensory information and results of interaction, and represent the relationship between sensor and behavior as stochastic behavior decision models. The robot advances the learning through making suggestions and questions for the user using the stochastic model. We investigate the feasibility of this method on obstacle avoidance tasks for mobile robots. Through experiments, we have confirmed that the mobile robot acquires avoidance behavior against change of environment through only several teaching. Also we have confirmed that the acquired models reflect the experience of interaction, therefore the model reflects personal preferences of teaching operation.

Key Words: Human-Robot Interaction, Bayesian Network, Personal Mobile Robot, Learning

1. は じ め に

人間の活動をサポートし、人間が活動する環境内において行動するパーソナルロボットでは、整備されていない未知環境における行動、人間からの不確実性を伴うタスク依頼など、行動に必要な知識を事前に持ち合わせておくことが不可能である状況下で確実に行動することが要求される。そのためには自律行動の獲得のための学習や適応が必要となるが、人間と共存するロボットというコンセプトから、ロボットを使用するユーザとのインタラクションを最大限に活用するべきである。まったく知識を持っていないロボットにとっては、ユーザからの教示は欠かすことのできない自律行動の獲得手法である。しかしある程度の自律性を持ったロボットにおいては、ユーザからの教示に完全に頼る必要はない。頼る部分は頼り、教示が必要でない部分に関しては、ユーザの手を煩わせることなく、過去に得られた教示情報を元にして自律性を確保すべきである。すなわちユーザとロボットがインタラクションを介して経験を重ねて行くうちに、段階的に自律性を高める手法がパーソナルロボットに求められる能力である。

Torrance [1] は移動ロボットにおいて自然言語で動作を指示

するシステムを構築したが、あらかじめ組み込まれた行動を対話的に切り替えるのみであり、センサ情報を元に行動を切り替える戦略を獲得する機能は持ち合わせていない。Shibata et al [2] [3] は人間が移動ロボットに乗り込み、ユーザがオンラインで行動と環境の見え方の関係を教示することで新しい行動を容易に獲得させるシステムを開発した。しかしながら Kuniyoshi [4] らの人間動作の観察に基づく教示手法と同じように、視覚情報の見え方と行動との関係が一对一に対応しており、環境の変動に対応不可能であったり、一般的な行動のモデルを獲得するまでには至っていない。

事情通ロボットプロジェクトでは、ユーザと対話を行うことのできる移動ロボットにおいて、未知環境の地図情報を獲得する枠組み [5] を提案しているが、人間とロボットとの対話を行動の獲得に直接結び付けてはいない。Asada et al [6] [7] は移動ロボットにおける強化学習において、センサ状態空間の構成法を提案している。この研究では過去にロボットが得たセンサデータの経験を効果的に利用しているが、ユーザとロボットのインタラクションは考慮されていない。Ishiguro et al [8] は、同じく移動ロボットでの強化学習における状態空間の構築の中で、学習の初期の段階において生じる探索空間の発散を抑えるために、人間が教示した行動を見本にして学習をスタートさせている。しかしながら学習が始まってからはロボットが内部状態を構築して行くだけで、ユーザとのかかわりは存在しない。

これらの研究に対して本研究では、ユーザに負荷をかけずにユーザの嗜好や意図をロボットの学習に反映させることを狙う。

原稿受付 2000 年 6 月 22 日

^{*1} 科学技術振興事業団, CREST

^{*2} 東京大学大学院情報理工学系研究科

^{*1} Japan Science and Technology Corporation, CREST

^{*2} Faculty of Eng., University of Tokyo

開発者がロボット内部に行動決定モデルをあらかじめ埋め込むのではなく、ユーザがロボットと対話を重ねながら段階的に環境に応じた行動決定モデルを獲得する仕組みが、パーソナルロボットにとって重要であると考えているからである。

本論文では具体例として、移動ロボットの障害物回避問題に焦点をあてる。障害物を回避する自律行動を獲得することに加え、その自律行動の獲得を効果的にするような、ロボットとのインタラクションをユーザに提供することを目的とする。

第2章では、統計的手法を用いて環境と行動の関係を表現する行動決定モデルの概略と利点について述べ、行動決定と対話制御の仕組みについて述べる。第3章では、ユーザとの対話に基づく行動決定モデルの獲得とそれを利用した対話制御の具体的手法について述べる。第4章で、移動ロボットにおける障害物の回避タスクへの応用について述べ、第5章で、実験結果を通して本手法の有効性を検証する。最後に第6章で考察と結論を述べる。

2. 経験の蓄積に基づく統計的行動決定モデル

2.1 統計的表現に基づく行動決定モデル

何の知識も持っていないロボットに新しい知識を教示する場合、一般的な教示方法では環境情報やロボットの内部状態に応じて、取るべき行動を一对一対応させるように教示していた。このように形成された行動決定モデルは、環境情報と行動の関係が細かく表現できるため、複雑な環境や状況に対応できる。しかし環境の変動などに影響を受けやすく、ロボットを使用するユーザが交代したり、センサの物理的特性が変化したりして獲得したモデルが使用できなくなると、また初めから教示をし直すことになり、いったん獲得した経験を再利用することが難しかった。また、人間の教示も完全である保障はなく、人間が教示ミスを行うと、そのままミスを反映するようなモデルが形成されてしまう。

これは、環境情報と取るべき行動の間の適切関係が一意的に存在することを前提として、固定的に表現されていたためと考えられる。この問題を回避するために本論文では、教示の結果を環境情報とダイレクトにバインドして記録するのではなく、統計を用いて確率的に表現する手法を導入する。統計を用いたセンサ行動間の表現手法の利点として以下のようなものがある。

- (1) ロボットと人間がインタラクションを重ねた経験から単純な統計計算を用いて、オンラインでの行動決定モデルの獲得が可能になる（具体的には第3.1節で述べる）。
- (2) ロボットが決定した行動の確からしさを確率値で表現可能であり、その確からしさを利用してユーザとのインタラクションを制御できる（具体的には第2.2節で述べる）。
- (3) 統計処理によりセンサノイズやユーザの操縦ミスなどの影響を抑え、未知の環境でもある程度の自律性を確保できる。

このようなモデルを構築するために、我々は Bayesian Network [9] を採用した。Bayesian Network は複数の事象間の因果関係を確率を用いて表現する推論モデルであり、実際に移動ロボットの地図獲得 [5] [10]、コンピュータのエラー診断 [11]、対話システム [12] など、幅広く応用されている手法である [13]。Bayesian Network では Fig. 1 のように事象をノードとして表

Fig. 1 Connection between behavior and sensor in Bayesian Network model

現し、事象間の確率的な関係をノード間のアークとして表現する。各事象はいくつかの離散状態を持ち、各状態を取り得る可能性を意味する確率変数が割り振られる。アークが意味する事象間の因果関係はその確率変数を使った条件付確率で表現される。また、ノードは大別して二種類から成る。一つは確率変数が観測される証拠ノード、もう一つは観測された証拠から推論を行う目的の事象を意味する仮説ノードである。

証拠ノード E の確率変数を $\mathbf{E} = \{e_1, e_2, \dots, e_n\}$ 、仮説ノード H の確率変数を $\mathbf{H} = \{h_1, h_2, \dots, h_m\}$ とする。仮説事象 h_j が観測されたときに証拠事象が e_i である事後確率を $P(e_i|h_j)$ としたとき、証拠ノード E と仮説ノード H の因果関係を次の行列

$$CPT_{E|H} \stackrel{\text{def}}{=} P(E|H) = \begin{pmatrix} P(e_1|h_1) & P(e_2|h_1) & \dots & P(e_n|h_1) \\ P(e_1|h_2) & P(e_2|h_2) & \dots & P(e_n|h_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(e_1|h_m) & P(e_2|h_m) & \dots & P(e_n|h_m) \end{pmatrix} \quad (1)$$

で表す。この行列を条件付き確率テーブル (Conditional Probabilities Table, 以下 CPT) と呼ぶ。

証拠事象が観測されている状態で、仮説事象の状態を推論する場合には、次式のベイズの定理より未観測事象の状態の確率を得ることができる。

$$BEL(\mathbf{H}) \stackrel{\text{def}}{=} P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{CPT_{E|H} \cdot P(H)}{P(E)} \quad (2)$$

2.2 統計的行動決定モデルに基づく自律行動と対話制御

ロボットにおける自律行動の決定モデルに使用する場合は、Fig. 1 のように証拠ノード E を外界の環境すなわちセンサ情報 S と対応させ、仮説ノード H をロボットに組み込まれている基本的行動モジュール B と結び付ける。すなわち、ユーザとの対話に基づく教示の経験から取るべき行動の事前確率 $P(B)$ と事後確率 $CPT_{S|B}$ を獲得し、自律行動を行う際にはセンサ情報の観測 $P(S)$ からとるべき行動 $P(B|S)$ を求める。

推論された結果であるベクトル $BEL(\mathbf{B})$ は確信度と呼ばれ、その名前の通り推論に対する確信の度合を意味し、これを用いて行動選択や対話制御を行う。例えば確信度が $BEL(\mathbf{B}) = (0.01, 0.97, 0.02)$ となった場合は、確率変数が意味する行動のうち2番目を選択する。一方 $BEL(\mathbf{B}) = (0, 0.5, 0.5)$ という確信度が得られた場合には2番目と3番目の行動を行うべき確

Table 1 Interaction processes and experience weights for each situation

Belief Degree of Reasoning[%]	The differences of user's operations from reasoning results		
	Same Case	Different Case	Case where the user doesn't operate
80~100	Obey user's operations	Execute reasoning results with point out user's mistake	Execute reasoning result
Autonomous Level 4	Storage with weight 0	Storage with weight 10 if the user corrects the reasoning result	Storage with weight 0
65~79	Obey user's operations	Wait user's operations with showing the reasoning result	Execute the reasoning result while showing deduced choices
Autonomous Level 3	Storage with weight 1	Storage with weight 5	Storage with weight 2
50~64	Obey user's operations	Obey user's operations with showing reasoning result	Execute the reasoning result while showing deduced choices
Autonomous Level 2	Storage with weight 3	Storage with weight 5	Storage with weight 5
0~49	Obey user's operations	Obey user's operations	Wait for user's operations while showing deduced choices
Autonomous Level 1	Storage with weight 5	Storage with weight 10	Storage with weight 10

率が等しい、という解釈になる。このような場合には、ユーザに対して「2と3、どちらの行動を取れば良いか教えて下さい」などと発話する。また、ロボットの推論と人間の教示が異なった場合にも推論の確信度を利用する。ロボットの確信度が高い場合はユーザのミスを指摘し、確信度が低い場合はユーザの指示に従う。どちらともいえない場合には、ロボットが選択肢をユーザに提示し指示を待つ。このように、単にユーザに質問を行うだけでなく、ロボットが自分の持つ選択肢の中から候補を絞ってユーザへ選択を委ねるという対話制御によって、Grice [14]の提案する協調的対話の原理を満たすような、ユーザに負荷をかけることのない、状況に応じた必要十分な量の対話を行うことができる。この対話制御および行動選択の戦略を **Table 1** の各枠上段に示す。

確率的手法の他にもニューラルネットワークのように非線形の対応関係を学習させる方法 [15] や、強化学習による行動獲得、分類木による行動決定 [16]、などが存在する。本手法では、ノードの状態にはユーザが理解しやすいシンボルが対応しており、それを用いてロボットが現在の状況をユーザに説明するのに適しているのに対して、ニューラルネットワークの場合は内部状態がユーザにとって理解することができず、インタラクションを前提としたロボットには不向きである。強化学習では報酬がシステムの中に固定的に埋め込まれているので、ユーザが後天的に与える指示に沿うような行動が獲得される保証はない。報酬関数とユーザの直観的な指示の間に距離があることが問題となる。しかも、学習に必要な行動の試行回数のオーダーが1,000~10,000に及ぶことが多く、実世界で人間と対話しながら行動するパーソナルロボットにとって適切ではない。また、Veloso et al [16] は、ユーザの教示データとセンサデータの関係を分類木(C4.5)を用いて表現しているが、分類木を用いた場合、センサノイズやユーザの教示ミスによって一つでも木の選択を誤ると適切な行動が出力されない危険性がある。本手法と同様に観測データの蓄積に基づく学習手法である Nearest Neighbor 法も確信度を利用した対話の制御を行うことは困難である。

それに対して統計的行動決定モデルは、過去の経験に基づく確信度を使用するため、実環境に対するフレキシブルな行動選択のみならず、ユーザとの対話に基づいて行動を学習する際に有利な点を持ち合わせている。

3. 統計的行動決定モデルの段階的獲得

Bayesian Network による行動決定モデルを獲得するには、

- (1) 先に述べた CPT、および事前確率の計算
 - (2) 各ノードにおいて連続値を離散的な状態に変換する際の戦略の決定（閾値や状態の数など）
 - (3) Bayesian Network の構造の決定
- の各操作が必要である。本論文では構造は固定とし、(1)、(2)に相当する手法について以下に述べる。

3.1 状況の重要度による重み付けを利用した CPT の計算

N 回の経験の中で、ユーザから行動 b_j が教示された回数が N_j 回であり、その教示が行われたうちで、センサ情報として s_i が観測された回数が n_{ij} だったとする。この場合もっとも単純な事後確率、事前確率の計算方法は次式ようになる。

$$P(s_i|b_j) = P(S = s_i|B = b_j) = \frac{n_{ij}}{N_j} \quad (3)$$

$$P(b_j) = P(B = b_j) = \frac{N_j}{N} \quad (4)$$

しかしこのような単純な累積頻度度数を用いた場合には問題が生じる。時々しか教示されない行動があった場合、事前確率 $P(B)$ が頻繁に教示された行動に偏ってしまい、滅多に教示されない行動を選択する確率が実際より小さい値になってしまう。

この原因は、観測されるセンサ情報パターンや、教示される行動パターンが、取り得る全パターン状態に対して均一な確率分布で観測されると仮定している所にある。人間の心理的な経験の蓄積においても、単に遭遇した頻度ではなく、状況の重要度で経験が重み付けされていると考えられる。本手法における「重要な状況」とは、ユーザの教示とロボットの推論が食い違った場合、確実な推論が行えない場合、の二つである。これらの状況は、ロボットにとって初めて遭遇した状況であり、自律行動獲得のために特に注目しなければならない特徴的なシーンである。そのため、通常の自律行動がうまく行われている状況に比べて重みをかけて経験に蓄積する。使用した状況別の重みを **Table 1** の各枠下段に示す。ユーザが教示を行わない自律行動時において、ロボットが確信度の高い推論を行った際には、十分に経験を重ねた状況とみなして重みを0として扱う。すなわち経験データに蓄えない。逆に確信度が50%以下の場合には今

Fig. 2 Weight for experience based on importance of situation

まで経験したことがない状況とみなして、重み 10 で経験に蓄えられる。

ユーザから行動 b_j が N_j 回教示されたうち、状況 c_k であった回数を $N_j^{(k)}$ とする。同様に行動 b_j の観測と同時にセンサ情報 s_i が観測された回数が n_{ij} だったうち、状況 c_k であった回数が $n_{ij}^{(k)}$ であったとする。状況 c_k の重みが w_k である場合に重み付けを考慮したことを変数に l を付けて表現し、

$$N_j' = \sum_k w_k N_j^{(k)} \quad (5)$$

$$n_{ij}' = \sum_k w_k n_{ij}^{(k)} \quad (6)$$

$$CPT_{S|B}' = P'(S = s_i | B = b_j) = \frac{n_{ij}'}{N_j'} \quad (7)$$

$$P'(B) = P'(B = b_j) = \frac{N_j'}{\sum_j N_j'} \quad (8)$$

とする。Fig. 2 にこの処理の模式図を示す。

確率値の推定には多くの観測データが必要であり、少量のデータしか存在しない場合には、ベータ分布や Dirichlet 分布などを用いて推定することが多い [17]。しかしながら、これらの分布を仮定したとしても、正しい確率値に収束させるために必要な観測データ数が減少するわけではない。状況に応じた重みを導入することで、ロボットは経験したことのない状況をより多く経験に蓄え、頻繁に遭遇する状況はあまり経験に蓄えない、という効果的な学習を行うことが可能となる。

3.2 過去の経験の内省による行動決定モデルの最適化

上記アルゴリズムでは、連続値のセンサデータをノードの状態値に離散化するために、本来は違うはずの状態が同じであると判断される、知覚の見せかけ問題 (perceptual aliasing) が起りやすい。この原因の一つはその場の状況に対して離散化のための閾値が適切でないことにある。適切な閾値は環境やユーザの教示の仕方によって変動する。すなわち、これは開発者があらかじめ設定できるパラメータではない。そこで CPT の計算と同様、蓄積された経験に応じて適応的に閾値を変動させるアプローチを取る。

過去に行われた教示行動の際に人間から与えられた教示行動と、その瞬間にロボットが推論した行動の相違を、行動決定モデルのエラーを表す指標として注目する。ある時刻 t においてロボットが推論した行動の確信度を $BEL_t(\mathbf{B})$ とし、そのとき

Fig. 3 Flowchart of proposed method

に人間が行った教示動作を $\mathbf{T}_t$ とする。これらをベクトルとして扱い、この時刻における行動決定モデルのエラー値を、二つのベクトルの各成分の差の絶対値の和

$$e_t = ||BEL_t(\mathbf{B}) - \mathbf{T}_t|| \quad (9)$$

で表現する。この値を教示が行われた全履歴に対して平均を取ったものを、行動決定モデルのエラー値

$$E = \frac{\sum_i^N e_i}{N} \quad (10)$$

とする。ただし N は履歴データの総数である。この値が小さければ、ユーザの教示を良く反映している行動決定モデルであると判断することができる。

この行動決定モデルのエラー値はセンサノードの閾値、取り得る状態の数などによって決まる。これらのパラメータ空間の中で最小のエラー値を求めることになるが、これについては探索空間が膨大になるため、遺伝的アルゴリズム (GA) を用いて探索を行う。すなわち、閾値などを遺伝子で表し、式 (10) で表される行動決定モデルのエラー値を適合度とする。

以上、3.1 節、3.2 節で述べた処理を繰り返し行うことで、統計的行動決定モデルを段階的に獲得して行く。Fig. 3 に全体的な処理の流れを示す。

4. 移動ロボットにおける障害物回避タスクへの応用

4.1 Bayesian Network による障害物回避タスクの表現

移動ロボットの障害物回避タスクへの応用を考えるため、センサノードとして距離センサを使用する。ロボットは Fig. 4 のように水平方向 1 ラインの距離を計測し、その結果を平滑化して 8 個の距離センサノードに対応させる。距離センサの値は整数値で、近い場合は 0 を、遠い場合は 255 を出力する。それらのデータを離散化し状態値に変換した上でノードに入力する。ここでは距離の確率変数として、 $\mathbf{S}_i = \{\text{Near}, \text{Far}\}$ の二つの状態を設ける。

またロボットの基本行動を表す行動ノードについては前進、ストップ、左折、右折の四つの行動要素を用意し、確率変数 $\mathbf{B} = \{\text{Forward}, \text{Stop}, \text{Left}, \text{Right}\}$ を割り振る。人間から与えられる指示はジョイスティックレバーによる指示と、音声言

Fig. 4 Distance sensors and environment condition

Fig. 5 Configuration of Bayesian Network

語による指示の二通りである。言語による指示の場合は各行動要素に対応する指示を受け付けるようにし、ジョイスティックレバーの場合には、ある閾値で離散化し、各行動要素と対応付ける。使用する Bayesian Network のモデルを Fig. 5 に示す。

複数のノードが階層的に結合している一般的な Bayesian Network における確信度の計算は複雑になるため、詳しくは文献 [9] にゆずるが、Fig. 5 において、センサノードの観測から、未知の状態である行動ノードを推論する際の確信度 $BEL(B)$ は次の式

$$BEL(B) = P(B|S) = \alpha \prod_{i=1}^n \lambda_i(B) \pi(B) \quad (11)$$

で表される。ただし α は正規化のための係数、 n はセンサノードの数、 $\pi(B)$ は行動ノードの事前確率、 $\lambda_i(B)$ は i 番目のセンサの観測結果から求められるベクトルで、 s_j という観測結果が得られた場合確率変数の j 番目の成分を 1、他の成分を 0 としたベクトル s_i を使って

$$\lambda_i(B) = CPT'_{S_i|B} s_i \quad (12)$$

と表される。例えば距離センサ S_i の入力 Near であれば、Near の確率が 1、Far の確率が 0 と考えて、 $s_i = (1, 0)$ というベクトルとなる。最終的に、式 (7) (8) を用いて、式 (11) は次の式になる。

$$BEL(B) = \alpha P'(B) \prod_{i=1}^8 CPT'_{S_i|B} s_i \quad (13)$$

4.2 仮想移動ロボットにおける実験システム

移動ロボットにおける障害物回避実験を行うために、Fig. 6 左のような仮想空間における移動ロボットと、Fig. 6 右のような仮想コックピットを作成した。仮想コックピットにはジョイスティック型の入力デバイスがあり、ロボットとの音声対話のためのマイクとスピーカも設置されている。ユーザはジョイ

Fig. 6 A virtual mobile robot and a user interface

Fig. 7 Three corridor environments

スティックを用いて基本的な教示行動を行い、ロボットからの質問や提案などには音声で答える。ロボットが自分の意図の提示や提案を行ったり、質問をする場合には、音声合成システムを通して音声発話を行う。

想定した仮想環境の鳥瞰図を Fig. 7 に示す。廊下環境を移動するロボットを想定しており、いくつかの障害物が配置されている。廊下の長さは 60、幅は 8 とした。ロボットは直径 2 の円柱で近似されている。センサは Fig. 4 のように、実機ロボットのカメラの画角と同じ 56° の中に均等に配置した。ユーザは左端の地点からロボットを操縦し、右端のつき当たりに到達して停止するような操縦を行う。

5. 実験結果

障害物回避の学習の進み方を Fig. 8 に示す。グラフの横軸はロボットが廊下環境を走行した回数を示しており、廊下の端から端までの走行で 1 回の走行とみなす。グラフの縦軸は確信度を示している。Fig. 8 (1) は直進すべきシーンであるが、学習を始めたばかりの初期の段階では、右折の行動が選択されている。また Fig. 8 (2), (3) はそれぞれ左折、右折を行うべきシーンであるが、2 回目の走行時においてロボットは直進しようとしており、ロボットは自律行動を解除しユーザの操縦に従っている。このグラフから、3 回の走行の後で、的確な自律行動が獲得されていることが分かる。

5.1 未知の環境への対応

本手法で用いている行動決定モデルは環境と行動の関係を統計的に表現しているため、未知の環境においても破綻することなく、ある程度の自律性をキープしながら行動を続行すること

Fig. 8 Progression of learning on a virtual mobile robot

ができる。

Fig. 7 の環境 A において回避戦略を学習した後に、未知の環境である B, C において自律行動を行わせた。Fig. 9 は各段階での操縦において、Table 1 の各状況が観測された度数分布を示している。この走行においては、5 回目に環境 B に移行し、12 回目に環境 C に移行している。図から分かる通り、最初の 1 回目のユーザの操縦と 10 回程度の質問行動のみで、異なる三つの環境についての自律行動を数回のオーダーの試行で獲得している。また Fig. 10 のように廊下に対して斜め方向からスタートさせたり、壁に近い位置からスタートさせた場合についても正常な軌跡を描くことがわかる。

ここで 3.1 節で述べた、状況の重要度に対する重み付けを行った場合と、行わなかった場合における自律走行の違いを Fig. 11 に示す。図は環境 C における自律走行の軌跡を示している。重み付けを行う場合は的確な回避行動が実現されたが、重み付けを行わない場合は障害物に衝突した。このことから、重み付け経験に基づく学習手法が優れていることがわかる。

5.2 実機ロボットでの検証

以上までに説明した仮想環境における移動ロボットでの実験を、実際の移動ロボットについても行った。移動ロボットの全体図と通路との関係を Fig. 12 に示す。このロボットはステレオカメラを用いて Fig. 4 で示した間隔で距離情報を取得する。

実際の移動ロボットにおける学習の進み具合を Fig. 13 に示す。ステレオカメラによる距離計測はノイズの影響が大きいですが、本研究と同じく数回の教示で学習が収束することを確認している。

Fig. 9 Distribution of each autonomous level

Fig. 10 Running result with change of initial condition

Fig. 11 Comparison weighting strategy with no weight strategy

Fig. 12 A real mobile robot

5.3 行動決定モデルのユーザ個人への適応

3.2 節で説明した状態空間の構成法は、ユーザの対話的教示の履歴に基づいている。そのため、異なる判断基準を持つユーザに対して、個人別に適応した行動決定モデルを獲得することができる。

まず、3.2 節の手法の単純な例を説明する。8 個のセンサノードすべてが Near, Far の二つの状態を取るとし、その閾値 T もすべてのノードで共通であるとする。環境 A, B, C, それぞれで走行した教示データに対する、閾値 T と行動決定モデルのエラー値 E の関係を Fig. 14 に示す。このグラフでは $T = 79$ のときにエラー値が最小値 0.87 となっている。グラフは単純な凹型をしており、この値から T が遠ざかるほど、エラー値 E

Fig. 13 Progression of learning on a real mobile robot

Fig. 14 Relation between distance threshold and error value

は増大している。エラー値が最小となる $T = 79$ を選んだ場合と、エラー値が大きい $T = 29$ および $T = 130$ の場合の自律走行結果を Fig. 15 に示す。エラー値が最小となる閾値 $T = 79$ の場合は的確に障害物を回避しているが、 $T = 29$ の場合は障害物に衝突し、 $T = 130$ の場合はゴール地点で停止できていない。このことからこのエラー値が、行動決定モデルの評価基準として適していることが分かる。

次に各センサノードに対して別々の閾値を設定することを考える。この場合は 3.2 節で述べた通り状態空間が膨大となるため、GA を用いて最小のエラー値を探索する。距離センサの状態値を Near, Middle, Far の三つの状態に離散化することを考え、そのために必要な二つの閾値を二つの遺伝子で表現する。すなわち 8 個のノードで計 16 個の遺伝子を使用する。個体数

Fig. 15 Running result for each threshold value

Fig. 16 Result of sensor space construction using GA

Fig. 17 Result for gentle operation and rough operation

100, 一点交差で交差率 0.6, 突然変異率 0.95%, エリート戦略の条件下において、第 62 世代目で Fig. 16 のような状態空間を形成した。この際の行動決定モデルのエラー値は 0.39 で、先の単純な例に比べてエラー値が非常に小さいことが分かる。

この状態空間の構成法がユーザ個人に適応可能であることを確認するために、ゆるやかに回避するように教示を行った場合と、障害物の直前まで回避しない教示を行った場合の二つの実験を行った。それぞれのユーザに対して獲得したモデルで自律走行を行った結果を Fig. 17 に示す。図から分かる通り同じ環境においてもユーザの教示を反映するような行動が生成されている。

6. おわりに

人間とロボットが対話を重ね、その経験から統計的処理を用いて自律行動の決定モデルを段階的に獲得する手法について述べた。その手法の実用性を検証するために移動ロボットの障害物回避タスクにおける実験を行った。統計的行動決定モデルの枠組では、ロボットは自分自身の行動決定に対する確信度を持つ。この確信度を用いることによって、ユーザに負荷をかけずに提案や質問を通じて経験を蓄えることが可能となり、他の学習

方法では困難となる, ユーザとの対話に基づく学習を実現した.

現段階の課題点として, センサ情報のある瞬間に対して使用しているだけで, 時系列データを用いて連続的なセンサ情報間の関係は考慮していない, という点がある. Bayesian Network を時系列方向に拡張する手法はすでに提案されており [18], 本研究においてもこのような拡張は非常に重要となってくると思われる. 一方, 移動ロボットの障害物回避というタスク以外にも, ユーザの使用する言葉とセンサ情報の間の関係を学習するタスク [19] [20] などに対して, 適切なネットワークモデルを組むことによって対応することができる.

あらかじめ報酬を埋め込む強化学習や, 開発者によって埋め込まれた特定条件下での行動モデルなどと異なり, 本手法はユーザの教示に基づいて行動決定モデルを獲得するため, 5.3 節で述べたようにユーザの教示の仕方に応じたモデルが獲得される. また, ユーザへの質問を頻繁に求めることを防ぐための, 確信度を利用した行動決定と対話制御が特徴となっている. このような, 対話に基づく学習, 段階的な自律行動の獲得, ユーザへの負荷をかけない対話制御, という技術要素の統合は日常空間にパーソナルロボットが普及するための必須条件であると考えている.

参 考 文 献

- [1] M.C. Torrance: "Natural Communication with Robots," Master's thesis, Massachusetts Institute of Technology, 1994.
- [2] T. Shibata, Y. Matsumoto, T. Kuwahara, M. Inaba and H. Inoue: "Development and integration of generic components for a teachable vision-based mobile robot," IEEE/ASME Transactions on Mechatronics, vol.1, no.3, pp.230-236, 1996.
- [3] 柴田, 松本, 稲葉, 井上: "人が搭乗しその場で行動を指示できるパーソナル視覚移動ロボット Hyper Scooter の開発", 日本ロボット学会誌, vol.14, no.8, pp.1138-1144, 1996.
- [4] Y. Kuniyoshi, M. Inaba and H. Inoue: "Learning by Watching: Extracting Reusable Task Knowledge from Visual Observation of Human Performance," IEEE Transaction on Robotics and Automation, vol.10, no.6, pp.799-822, 1994.
- [5] H. Asoh, Y. Motomura, I. Hara, S. Akaho, S. Hayamizu and T. Matsui: "Acquiring a Probabilistic Map with Dialogue-Based Learning," In Proceedings of ROBOLEARN-96, 1996.
- [6] M. Asada, S. Noda and K. Hosoda: Action-Based Sensor Space Categorization for Robot Learning," In Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS'96), pp.1502-1509, 1996.
- [7] 高橋, 浅田: "実ロボットによる行動学習のための状態空間の漸近的構成", 日本ロボット学会誌, vol.17, no.1, pp.118-124, 1999.
- [8] H. Ishiguro, R. Sato and T. Ishida: "Robot Oriented State Space Construction," In Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS'96), pp.1496-1501, 1996.
- [9] J. Pearl: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.
- [10] K. Basye, T. Dean, J. Kirman and M. Lejter: "A Decision-Theoretic Approach to Planning Perception and Control," IEEE Expert, vol.7, no.4, pp.58-65, 1992.
- [11] D. Heckerman, A. Mamdani and M.P. Wellman: "Real-World Applications of Bayesian Networks," Communicaiton of the ACM, vol.38, no.3, pp.24-57, March 1995.
- [12] T. Akiba and H. Tanaka: "A Bayesian Approach for User Modeling in Dialogue Systems," Technical report, Dept. of Computer Science, Tokyo Institute of Technology, 1994.
- [13] E. Charniak: "Bayesian Networks without Tears," AI Magazine, vol.Winter, pp.50-63, 1991.
- [14] H.P. Grice: Logic and Conversation, Peter Cole and Jerry L.Morgan, editors, Syntax and Semantics vol.3: Speech acts. Academic Press, pp.41-58, 1975.
- [15] D.A. Pomerleau: "ALVINN: An autonomous land vehicle in a neural network," Technical Report CMU-CS-89-107, CMU, 1989.
- [16] M. Veloso, W. Uther, M. Fujita, M. Asada and H. Kitano: "Playing Soccer with Legged Robots," In Proceedings of the 1998 IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS'98), pp.437-442, Oct. 1998.
- [17] D. Heckerman, D. Geiger and D. Chickering: "Learning bayesian networks: The combination of knowledge and statistical data," Machine Learning, vol.20, pp.197-243, 1995.
- [18] J. Forbes, T. Huang, K. Kanazawa and S. Russell: "The BATmobile: Towards a Bayesian Automated Taxi," In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95), pp. 1878-1885, 1995.
- [19] 稲邑, 稲葉, 井上: "個人に適応した語彙を獲得するロボットとの自然言語対話処理-ヒューマンロボットインタラクションシステム PEXIS の機能その 3-", 第 5 回ロボティクスシンポジウム予稿集, pp.146-151, 2000.
- [20] 稲邑, 稲葉, 井上: "PEXIS:統計的経験表現に基づくパーソナルロボットとの適応的インタラクションシステム", 電子情報通信学会誌, vol.J84-D-I, no.6, pp.867-877, 2001.

稲邑哲也 (Tetsunari Inamura)

1973 年 1 月 23 日生. 1995 年東京大学工学部卒業. 1997 年度から 1999 年度まで日本学術振興会特別研究員. 2000 年同大学大学院工学系研究科情報工学専攻博士課程修了. 博士 (工学). 同年科学技術振興事業団 CREST 研究員, 現在に至る. ヒューマンロボットインタラクション, 脳の情報処理機構の開発等の研究に従事. 人工知能学会及び電子情報通信学会会員.

(日本ロボット学会正会員)

井上博允 (Hirochika Inoue)

1942 年 7 月生. 1965 年東京大学工学部卒業. 1970 年同大学大学院博士課程修了. 工学博士. 同年電子技術総合研究所入所. 知能ロボットの研究開発に従事. 1977 年東京大学工学部機械工学科助教授. 1984 年教授. 現在, 機械情報工学科教授. ロボット全般, 人工知能, 情報システム工学の研究と教育に従事.

(日本ロボット学会正会員)

稲葉雅幸 (Masayuki Inaba)

1958 年 5 月 23 日生. 1986 年東京大学大学院工学系研究科情報工学専門課程博士課程修了. 工学博士. 1986 年東京大学講師, 1989 年助教授, 2000 年東京大学教授, 大学院工学系研究科機械情報工学専攻. 現在大学院情報学環・学際情報学府所属. 日本機械学会, 計測自動制御学会, 情報処理学会, 人工知能学会, ソフトウェア学会等会員.

(日本ロボット学会正会員)