(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2013-205890
(P2013-205890A)
(43)公開日　平成25年10月7日(2013.10.7)

(51)Int.Cl.		F I		テーマコード（参考）
G06N 3/00	(2006.01)	G06N 3/00	560A	
G06N 5/04	(2006.01)	G06N 5/04	550J	
G06F 17/30	(2006.01)	G06F 17/30	210D	

審査請求 未請求　請求項の数 6　OL　（全 15 頁）

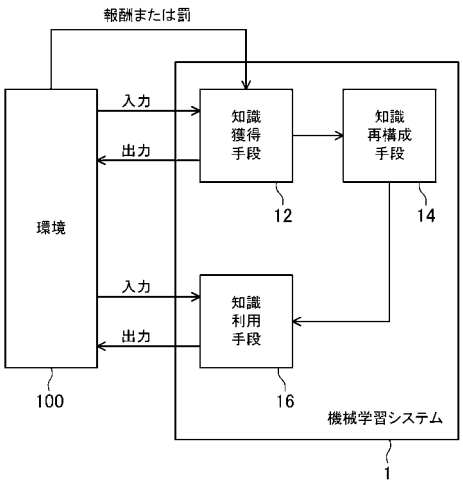
(21)出願番号　特願2012-71205(P2012-71205)
(22)出願日　平成24年3月27日(2012.3.27)

特許法第３０条第１項適用申請有り　平成２３年１１月２１日に公益社団法人　計測自動制御学会　システム・情報部門の学術講演会２０１１　講演論文集にて発表

(71)出願人　504136568
国立大学法人広島大学
広島県東広島市鏡山１丁目３番２号
(74)代理人　110001427
特許業務法人前田特許事務所
(72)発明者　保田　俊行
広島県東広島市鏡山一丁目４番１号　国立大学法人広島大学大学院工学研究院内
(72)発明者　大倉　和博
広島県東広島市鏡山一丁目４番１号　国立大学法人広島大学大学院工学研究院内

(54)【発明の名称】機械学習システムおよび機械学習方法

(57)【要約】
【課題】パラメトリック表現された状態空間とノンパラメトリック表現された状態空間とを適応的に選択する。
【解決手段】機械学習システム（１）は、入力および当該入力に対する出力に対して与えられる報酬または罰に基づいて強化学習を行って、パラメトリック表現されたクラス集合を生成する知識獲得手段（１２）と、パラメトリック表現されたクラス集合の生成に使用された学習済み入力に基づいて、ノンパラメトリック表現されたクラス集合を生成する知識再構成手段（１４）と、未知の入力がノンパラメトリック表現されたどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をする知識利用手段（１６）とを備えている。知識再構成手段（１４）は、学習済み入力の個数が所定数よりも多く、かつ、パラメトリック表現された各クラスの分散が所定値よりも小さいとき、ノンパラメトリック表現されたクラス集合を生成する。
【選択図】図１



【特許請求の範囲】
【請求項１】
入力が状態空間におけるどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をし、入出力を繰り返すことで環境に適応した知識を獲得する機械学習システムであって、
入力および当該入力に対する出力に対して与えられる報酬または罰に基づいて強化学習を行って、パラメトリック表現されたクラス集合を生成する知識獲得手段と、
前記パラメトリック表現されたクラス集合の生成に使用された学習済み入力に基づいて、ノンパラメトリック表現されたクラス集合を生成する知識再構成手段と、
未知の入力が前記ノンパラメトリック表現されたどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をする知識利用手段とを備え、
前記知識再構成手段は、前記学習済み入力の個数が所定数よりも多く、かつ、前記パラメトリック表現された各クラスの分散が所定値よりも小さいとき、前記ノンパラメトリック表現されたクラス集合を生成する
ことを特徴とする機械学習システム。
【請求項２】
請求項１に記載の機械学習システムにおいて、
前記パラメトリック表現された各クラスが多変量の正規確率分布であり、
前記知識獲得手段は、ベイズ判別法に従って、入力が前記パラメトリック表現されたどのクラスに属するかクラス判別を行う
ことを特徴とする機械学習システム。
【請求項３】
請求項１および２のいずれか一つに記載の機械学習システムにおいて、
前記知識再構成手段は、サポートベクターマシンを用いて前記学習済み入力を線形分離して、前記ノンパラメトリック表現されたクラス集合を生成する
ことを特徴とする機械学習システム。
【請求項４】
入力が状態空間におけるどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をし、入出力を繰り返すことで環境に適応した知識を獲得する機械学習方法であって、
入力および当該入力に対する出力に対して与えられる報酬または罰に基づいて強化学習を行って、パラメトリック表現されたクラス集合を生成する第１のステップと、
前記パラメトリック表現されたクラス集合の生成に使用された学習済み入力に基づいて、ノンパラメトリック表現されたクラス集合を生成する第２のステップと、
未知の入力が前記ノンパラメトリック表現されたどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をする第３のステップとを備え、
前記第２のステップでは、前記学習済み入力の個数が所定数よりも多く、かつ、前記パラメトリック表現された各クラスの分散が所定値よりも小さいとき、前記ノンパラメトリック表現されたクラス集合が生成される
ことを特徴とする機械学習方法。
【請求項５】
請求項４に記載の機械学習方法において、
前記パラメトリック表現されたクラスが多変量の正規確率分布であり、
前記第１のステップでは、ベイズ判別法に従って、入力が前記パラメトリック表現されたどのクラスに属するかクラス判別が行われる
ことを特徴とする機械学習方法。
【請求項６】
請求項４および５のいずれか一つに記載の機械学習方法において、
前記第２のステップでは、サポートベクターマシンを用いて前記学習済み入力が線形分離され、前記ノンパラメトリック表現されたクラス集合が生成される

ことを特徴とする機械学習方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、機械学習に関し、特に、強化学習による知識獲得の頑健性向上に関する。
【背景技術】
【０００２】
システムを制御する場合、一般的にはモデル化に基づくトップダウン的アプローチがとられる。しかし、システムの大規模化などの要因により制御が困難になるということも考えられる。一方、ボトムアップアプローチではシステムの構成要素を知能化することで系全体としての合目的的な入出力関係の獲得が可能である。その中の一つに強化学習法がある。強化学習法は、目標状態を与えるのみでそこに至る入出力の系列を自律的に構築できるという実装の容易さから、さまざまなシステムへの応用が期待される。
【０００３】
強化学習法の従来の枠組みでは、離散的な状態・行動空間における写像関係の構築を対象としている。ここで、学習性能はこの状態・行動空間の離散化具合に大きく影響されるが、現在のところそのための設計指針は存在していない。この問題は、連続空間において動作する多くの実システムでは重大な課題である。本願発明者らはこの状態・行動空間の設計問題に対する手法として、強化学習を機能拡張したBayesian-discrimination-function-based Reinforcement Learning（ＢＲＬ）を研究・開発してきた。ＢＲＬは、連続な状態・行動空間を自律的に分割する機能を持つ。さらには、従来型強化学習はマルコフ環境において学習収束が保証されているのみであるが、ＢＲＬは学習過程で分割具合を適応的に更新可能であるために動的環境でも学習可能であるという特徴を持つ。これまで、本願発明者らは、実システムとしてロボット、特に複数のロボットで構成されるマルチロボットシステム（Multi-Robot Systems：ＭＲＳ）を取り上げ、自律移動ロボット群やアーム型ロボット群による協調問題においてＢＲＬの有効性を示してきた。
【０００４】
ところが、その後の追加実験において、ＢＲＬでは行動獲得後にさらに学習を続けると、徐々にその頑健性が損なわれる場合があることが観察された。これは、タスク達成に寄与しないルールは削除され、寄与するルールのみが強化されてルール集合に残ることが原因である。すなわち、ＢＲＬでは環境に特化したルール集合となる結果、過学習状態となるためにシステムが不安定になる。そこで、近年、本願発明者らはパターン認識手法の一つであるSupport Vector Machine（ＳＶＭ）の高い識別性能に着目し、ＳＶＭによるルール判別がＢＲＬの過学習抑制に有効であることを明らかにした（例えば、非特許文献１参照）。
【先行技術文献】
【非特許文献】
【０００５】
【非特許文献１】J. Sakanoue, T. Yasuda, and K. Ohkura, "Preservation and Application of Acquired Knowledge Using Instance-Based Reinforcement Learning," Joint 5th International Conference on Soft Computing and Intelligent Systems and 10th International Symposium on advanced Intelligent Systems, 2010, pp.576-581
【発明の概要】
【発明が解決しようとする課題】
【０００６】
ＢＲＬの過学習抑制にＳＶＭが有効であることは実証できたものの、具体的にＢＲＬのルール判別にどのようにＳＶＭを用いるかについてはまだ提案できていない。かかる問題に鑑み、本発明は、機械学習システムにおいて、パラメトリック表現された状態空間とノンパラメトリック表現された状態空間とを適応的に選択する手法を提供することを目的とする。

【課題を解決するための手段】
【０００７】
本発明の一局面に従った機械学習システムは、入力が状態空間におけるどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をし、入出力を繰り返すことで環境に適応した知識を獲得する機械学習システムであって、入力および当該入力に対する出力に対して与えられる報酬または罰に基づいて強化学習を行って、パラメトリック表現されたクラス集合を生成する知識獲得手段と、前記パラメトリック表現されたクラス集合の生成に使用された学習済み入力に基づいて、ノンパラメトリック表現されたクラス集合を生成する知識再構成手段と、未知の入力が前記ノンパラメトリック表現されたどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をする知識利用手段とを備えている。前記知識再構成手段は、前記学習済み入力の個数が所定数よりも多く、かつ、前記パラメトリック表現された各クラスの分散が所定値よりも小さいとき、前記ノンパラメトリック表現されたクラス集合を生成する。
【０００８】
これによると、知識獲得手段による強化学習が十分に進んだところで、知識獲得手段において生成されたパラメトリック表現されたクラス集合が知識再構成手段によってノンパラメトリック表現されたクラス集合に再構成され、当該ノンパラメトリック表現されたクラス集合を用いて知識利用手段によって未知の入力のクラス判別が行われる。
【０００９】
具体的には、前記パラメトリック表現された各クラスが多変量の正規確率分布であり、前記知識獲得手段は、ベイズ判別法に従って、入力が前記パラメトリック表現されたどのクラスに属するかクラス判別を行う。
【００１０】
また、具体的には、前記知識再構成手段は、ＳＶＭを用いて前記学習済み入力を線形分離して、前記ノンパラメトリック表現されたクラス集合を生成する。
【発明の効果】
【００１１】
本発明によると、機械学習システムにおいて、パラメトリック表現された状態空間とノンパラメトリック表現された状態空間とが適応的に選択され、機械学習システムの頑健性が向上する。これにより、機械学習システムが環境変化にも柔軟に対応することができるようになる。
【図面の簡単な説明】
【００１２】
【図１】本発明の一実施形態に係る機械学習システムの機能ブロック図
【図２】図１の機械学習システムによる知識獲得および利用のフローチャート
【図３】計算機実験の実験環境を示す模式図
【図４】各ＳＶＭによるタスク達成率を示すグラフ
【図５】知識獲得までに要したエピソード数を示すグラフ
【発明を実施するための形態】
【００１３】
以下、図面を参照しながら本発明を実施するための形態について説明する。なお、本発明は、以下の実施形態に限定されるものではない。
【００１４】
（機械学習システムの実施形態）
図１は、本発明の一実施形態に係る機械学習システムの機能ブロック図である。本実施形態に係る機械学習システム１は、入力が状態空間におけるどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をし、入出力を繰り返すことで環境１００に適応した知識を獲得するものである。機械学習システム１は、例えばＭＲＳを構成する各ロボットなどに適用可能である。
【００１５】

機械学習システム１は、知識獲得手段１２、知識再構成手段１４、および知識利用手段１６を備えている。これら各手段は電子デバイスなどのハードウェアとして実現してもよいし、コンピュータで実行されるソフトウェアモジュールとして実現することもできる。以下、各手段について詳細に説明する。

【００１６】

（知識獲得手段１２の詳細説明）

知識獲得手段１２は、環境１００からの入力および当該入力に対する出力に対して与えられる報酬または罰に基づいて強化学習を行って、パラメトリック表現されたクラス集合を生成する。例えば、知識獲得手段１２はＢＲＬによって強化学習を行う。すなわち、パラメトリック表現された各クラスは多変量の正規確率分布であり、知識獲得手段１２は、ベイズ判別法に従って、入力がパラメトリック表現されたどのクラスに属するかクラス判別を行う。

【００１７】

≪ＢＲＬ≫

ＢＲＬでは統計的にパターン分類を行うベイズ判別法を用いて入力ｘがｋ番目（ただし、ｋは１からＮまでの整数である。）のクラスC_kに分類されるかを識別する。ベイズ判別法は、識別対象のクラス$C=\{C_k\}_{k=1}^{K}$および各クラスの事前確率Ｐ（C_k）と確率分布ｐ（ｘ｜C_k）が既知の場合、入力ｘが観測されたときの各クラスC_kの事後確率Ｐ（C_k｜ｘ）をベイズの公式から求め、事後確率最大となるクラスに入力を識別する方法である（数式（１）参照）。ＢＲＬでは、（１）クラスの追加と削除、（２）確率分布モデルのパラメータ更新によって観測データから環境の確率モデルをリアルタイムに更新し、状態空間の分割を行う。

【００１８】

【数１】

$$P(C_k|x)=\frac{P(C_k)p(x|C_k)}{\sum_{k=1}^{K}P(C_k)p(x|C_k)} \qquad (1)$$

【００１９】

≪ルール構成≫

各クラスをガウス分布によって表現し、各クラスの確率分布を表すパラメータとそのときの出力をif-then形式で記述したルールとして知識獲得手段１２に記憶する。これ以降、クラスとルールを同義として扱う。ルール集合Ｒはルール$rl\in R$により構成され、各ルールは次式で記述される。

【００２０】

【数２】

$$rl:=\langle v,\Sigma,f,u,\Phi,a\rangle \qquad (2)$$

【００２１】

各ルールｒｌは特徴ベクトル$v=\{v_1,\cdots,v_{nd}\}^T$、共分散行列Σ、クラスの事前確率ｆ、クラスの信頼性を表す有効度ｕ、各クラスで観測されたセンサ入力の集合$\Phi=\{\Phi_1,\cdots,\Phi_{ns}\}^T$、そして、動作$a=\{a_1,\cdots,a_{na}\}^T$より構成されている。ただし、$n_d$は入力空間の次元数、$n_a$は出力空間の次元数、$n_s$は各クラスが記憶しているサンプルデータを表す。学習初期、状態空間にはクラスは存在せず、機械学習システム１が実際に観測した入出力をもとに状態空間にクラスを追加し、状態空間をガウス分布で覆っていく。

【００２２】

≪動作選択≫

以下にＢＲＬの行動選択の概要を示す。

【００２３】

・知覚した入力を、ベイズ判別法によりどのクラスに属するかの判別を行う。

【００２４】

・既存のルールに属さない場合、ランダム行動を出力し、罰を受けなければ新たなルールを作成する。

【００２５】

・既存のルールに属す場合、そのルール行動を出力する。

【００２６】

入力に対する各ルールの事後確率をベイズの公式から求め、事後確率最大のルールに記述されている出力を実行する。ここでは、まず事後確率の負の対数を取り、誤って識別する確率g_iが最小となるルールを勝者ルールrl_wとする。

【００２７】

【数３】

$$g_w = \min_i \{g_i\} \qquad i \in [0, n_{rl}] \tag{3}$$

【００２８】

【数４】

$$g_i = -log\{f_i \cdot p(X|C_i)\} \tag{4}$$

【００２９】

このとき、事後確率が非常に小さいルールが選択されるのは適切でないと考え、事後確率に閾値P_{th}を設ける。そして、それをもとに計算される閾値$g_{th} = -\log\{f_0 \cdot P_{th}\}$によって$rl_w$の動作を実行するかどうか判断する。具体的には、$g_w < g_{th}$の場合、$rl_w$の動作$A_w$を実行する。$g_w \geqq g_{th}$の場合、ランダムに動作を実行する。なお、$f_0$および$P_{th}$は定数である。

【００３０】

≪有効度の更新≫

Profit SharingとBucket Brigade的戦略により報酬を過去に遡って伝播させる。その他、ループ行動を防ぐために選択されたルールに課すコスト、報酬獲得に寄与しないルールを削除してメモリ消費量を抑えるためにタスク達成時に全ルールに作用させる消散がある。

【００３１】

≪パラメータの更新≫

各ルールは入力をもとに確率分布のパラメータをオンラインで更新していく。リアルタイムに更新することで環境やシステム変動に対する迅速な対応が期待できる。その反面、ノイズや一時的な入力の偏りに影響を受けやすいため何らかの対処が必要となる。ＢＲＬでは、区間推定法を用いたパラメータ更新によりこの問題を解決する。区間推定法は、確率分布のパラメータがある区間に入る確率を設定した確率以上になるように保証する手法であり、サンプルデータが増大するにつれて推定精度が上がる。そのため、観測データの増加に伴ってより信頼性の高いパラメータ推定が期待できる。

【００３２】

【数５】

$$v_j \leftarrow v_j + \alpha(\bar{x}_j - v_j) \tag{5}$$

【００３３】
【数６】

$$\sigma_j^2 \leftarrow \sigma_j^2 + \alpha^2\left(s_j^2 - \sigma_j^2\right) \qquad (6)$$

【００３４】
【数７】

$$f \leftarrow \begin{cases} f + \beta(1-f) & (if\ P \geq 0) \\ (1-\beta)f & (otherwise) \end{cases} \qquad (7)$$

【００３５】
ここで、ｖはｒｌ$_w$の平均、σ^{2}はｒｌ$_w$の分散、α，βは区間［０，１］の定数、ｊは入力ベクトルにおける各次元、ｘバーはサンプル入力の平均、ｓ^{2}はサンプル入力の分散、Ｐは報酬である。
【００３６】
≪既存ルールのパラメータに基づく行動空間の適応的探索≫
常に行動空間をランダムに探索するのは非効率であるという観点から、知識獲得がある程度行われた状況では幅広く行動空間を探索するよりも既存のルールの近傍を探索して行動の調整を行うことが有効であると考える。そこで、ランダム探索をするための閾値Ｐ$_{th}$の他に新たに閾値Ｐ’$_{th}$を設定（Ｐ’$_{th}$＜Ｐ$_{th}$）し、ｇ$_{th}$≦ｇ$_w$＜ｇ’$_{th}$の場合はその間にあるルールのパラメータを参照して新しいルールパラメータを決定する。つまり、行動選択を以下のように変更する。
【００３７】
・ｇ$_w$＜ｇ$_{th}$の場合、ｒｌ$_w$の動作Ａ$_w$を実行する。
【００３８】
・ｇ$_{th}$≦ｇ$_w$＜ｇ’$_{th}$の場合、この間にあるルールをもとに動作を生成する。
【００３９】
・ｇ$_w$≧ｇ’$_{th}$の場合、ランダムに動作を実行する。
【００４０】
≪ｇ$_{th}$≦ｇ$_w$＜ｇ’$_{th}$の行動≫
この範囲には、ｒｌ$_w$以外にも複数のルールが含まれる場合がある。これらのルールはその状況化での選択確率としては大きな差はないものの、それまでの学習過程におけるタスク達成への貢献度に従って有効度が異なる。そのため、新しいルールの動作Ａ’は、この範囲に含まれるルールの有効度に基づく加重平均により求める。
【００４１】
【数８】

$$A' = \sum_{l=1}^{n_r}\left(\frac{u_l}{\sum_{k=1}^{n_r} u_k} \cdot A_l\right) + N(0,\sigma_{act}) \qquad (8)$$

【００４２】
ｎ$_r$はこの範囲に含まれるルール数であり、Ｎ（０，σ）は平均０・標準偏差σの正規分布を用いたノイズである。ノイズを付加することで、ｒｌ$_w$以外のルールがない場合であっても、Ａ$_w$の近傍を探索することができる。
【００４３】
（知識再構成手段１４および知識利用手段１６の詳細説明）
知識再構成手段１４は、知識獲得手段１２においてパラメトリック表現されたクラス集合の生成に使用された学習済み入力に基づいて、ノンパラメトリック表現されたクラス集

合を生成する。例えば、知識再構成手段１４はＳＶＭを用いて、知識獲得手段１２における学習済み入力を線形分離して、ノンパラメトリック表現されたクラス集合を生成する。知識利用手段１６は、環境１００からの未知の入力がノンパラメトリック表現されたどのクラスに属するかクラス判別を行って当該判別結果に応じた出力をする。

【００４４】

≪ＳＶＭによる知識再構成と利用≫

ＳＶＭは高い識別性能を持つとさまざまな分野で示されている。そこで、ＢＲＬの獲得知識のデータをＳＶＭにより識別することで、より正確な行動決定が可能になると期待できる。

【００４５】

本願発明者らは、ＢＲＬによりタスク達成に有効なルールを獲得できずサンプルデータが不十分な状態で、ＳＶＭによる知識利用を行うことは有効ではないと過去に示した（非特許文献１参照）。そこで、知識獲得手段１２における学習が十分に進み、機械学習システム１の行動が安定し始めたタイミングで知識再構成手段１４を動作させて獲得知識を再構成するため指標を設ける。

【００４６】

まず、ＳＶＭによる知識利用の前提として、サンプルデータが十分に多い、すなわち、知識獲得手段１２においてパラメトリック表現されたクラス集合の生成に使用された学習済み入力の個数が十分に多い必要がある。したがって、学習済み入力の個数が所定数よりも多いことを指標に設定する。

【００４７】

また、サンプルデータ数が十分に多くても、パラメトリック表現されたクラスの中に分散が大きいクラスが存在すると、ＳＶＭによる十分な識別精度が得られないおそれがある。そこで、パラメトリック表現された各クラス（ルール）の分散が所定値よりも小さいことも指標に設定する。例えば、この指標を各ルールの構成要素である共分散行列Σにより設定する。Σの値は状態空間においてルールの範囲を表す。行動が収束し始めるとΣの値が収束していく。よって、知識再構成手段１４は、例えば、エピソード毎の各ルールの｜Σ｜の平均（Σ_{eps}）を計算し、前エピソードの｜Σ｜の平均（Σ_{eps-1}）との差が閾値以下の場合、ＳＶＭによる知識再構成を行って、ノンパラメトリック表現されたクラス集合を生成する。そして、知識利用手段１４は、当該ノンパラメトリック表現されたクラス集合を用いてクラス判別、つまりＳＶＭによるルール選択を行う。

【００４８】

【数９】

$$\Sigma'_{eps} = \left| \Sigma_{eps} - \Sigma_{eps-1} \right| \qquad (9)$$

【００４９】

また、ルール判別可とする範囲をＢＲＬの場合よりも広くするために新たな閾値P_s（$P_s > P_{th}$）を設定する。閾値を広げることで新ルールの生成を抑制し、振る舞いの不安定化を防ぐことが期待される。以下に、機械学習システム１の行動選択の概要を示す。また、図２に、機械学習システム１による知識獲得および利用の概要を示す。

【００５０】

・知覚した入力をもとに、ＢＲＬにより知識探索を行うかＳＶＭにより知識利用を行うかを、Σ'_{eps}の値と閾値との比較により決定する。

【００５１】

・ＢＲＬにより知識探索を行う場合、ランダム行動を出力し、罰を受けなければ新たなルールを作成する。

【００５２】

・ＳＶＭにより知識利用を行う場合、ＳＶＭにより状態空間の再分割を行い、再分割し

た状態空間により判別されたルールの行動を出力する。

【００５３】

なお、数式（９）は指標の一例であり本発明はこれに限定されない。例えば、Σ'_{eps}の移動平均や加重平均を利用してもよい。

【００５４】

≪多クラス分類ＳＶMs≫

ＳＶMは基本的には２クラスの識別問題を対象にして定式化されている。しかし、２クラスの判別モデルを組み合わせることで多クラス分類を可能にしている。組み合わせ方としてOne-versus-AllとOne-versus-Oneという２種類を取り上げる。One-versus-Allとは、全クラスに対して、ある一つのクラスとそれ以外のクラスに分ける識別平面を作成し、これらの識別平面のうち最も高い判別値を返すクラスを出力するという方法である。ｎクラスの問題の場合、識別平面の数はｎとなる。一方、One-versus-Oneとは、各クラス毎に対となる識別平面を作成し、多数決により出力を決定する方法である。識別平面の数はｎ（ｎ－１）／２となる。知識利用手段１４に用いるＳＶMに、この２種類の多クラス分類方式を導入する。

【００５５】

≪ＳＶMに用いるカーネル関数≫

知識再構成手段１４で使用するＳＶMには高次元空間への写像により非線形分離を可能にするカーネルトリックを用いる。カーネルは以下に示す線形カーネルK_{line}、多項式カーネルK_{poly}、ＲＢＦカーネルK_{RBF}、シグモイドカーネルK_{sig}を使用する。ｕはサポートベクトル、ｖは識別する特徴ベクトルを表す。

【００５６】

【数１０】

$$K_{line} = \boldsymbol{uv} \tag{10}$$

【００５７】

【数１１】

$$K_{poly} = (\gamma \boldsymbol{uv} + c)^d \tag{11}$$

【００５８】

【数１２】

$$K_{RBF} = exp\left(-\gamma |\boldsymbol{u} - \boldsymbol{v}|^2\right) \tag{12}$$

【００５９】

【数１３】

$$K_{sig} = tah(\gamma \boldsymbol{uv} + c) \tag{13}$$

【００６０】

（計算機実験）

ＢＲＬによる知識獲得のタスクとして、二台のロボットによるピアノ運び問題を取り上げ、計算機実験を行う。ピアノ運び問題とは単体では搬送不可能な長尺物をロボットが協調しゴールまで搬送する問題である。狭い通路を通行するためには、二台のロボットが協

調し、フォーメーションを形成しなければならない。

【００６１】

実験環境を図３（ａ）に示す。フィールドは四方を壁で囲まれており、初期位置からゴールラインまで移動するとタスク達成となる。ロボットは差動駆動型を用い、二輪の駆動輪を持つ。各ロボットは全方位カメラにより他のロボットの状態を知覚し、ロボット間で通信を行わない。ロボットの一度の意思決定を１ステップとし、タスクを達成するか４００ステップ経過した時点でエピソードを更新し、ロボットを初期位置に戻す。学習成功の定義は、２０エピソード連続でタスクを達成したときとする。５００エピソード経過までを１試行とし、１００試行繰り返す。なお、シミュレーション環境はオープンソース三次元物理エンジンＯＤＥ（Open Dynamic Engine）により作成している。

【００６２】

≪実験１≫

ＢＲＬのみを用いて学習を行う。学習が成功した試行には、図３（ｂ）に示したような環境変化を行い、ＳＶＭによる知識利用の効果を検証するためＳＶＭのみで行動選択を行う。環境変化は通路の幅が狭くなることであり、タスクの難易度が上昇するため、獲得知識を状態に応じて正確に判別することが求められる。ここで使用する多クラス分類ＳＶＭｓは、２種類の多クラス分類方式と４種類のカーネル関数の組み合わせの８パターンを使用する。それぞれの組み合わせを表１で示すＡ～Ｈまでの記号で表す。学習が成功した試行中、環境変化のタスクを達成できた試行の成功率を観察する。

【００６３】

【表１】

	1-All	1-1
$K_{line}[(C,\gamma)=(2^3,2^{-2})]$	A	B
$K_{poly}[(C,\gamma,d,c)=(2^5,2^{-3},2,30)]$	C	D
$K_{RBF}[(C,\gamma)=(2^4,2^0)]$	E	F
$K_{sig}[(C,\gamma,c)=(2^4,2^{-4},0)]$	G	H

【００６４】

≪実験２≫

実験１で成功率の高かった組み合わせパターンについて、機械学習システム１を用いて学習を行う。学習するのに要したエピソード数の推移を調べ、機械学習システム１の行動獲得に対する有効性を検証する。パラメータチューニングにより、ＳＶＭを使用するΣ’$_{eps}$の閾値を０．０２と規定した。

【００６５】

≪機械学習システム１の設定≫

入力は、Ｉ＝｛r_0，$\cos\theta_0$，$\sin\theta_0$，r_1、$\cos\theta_1$，$\sin\theta_1$，r_2，$\cos\theta_2$，$\sin\theta_2$，$\cos\theta_3$，$\sin\theta_3$｝の１１次元である。ｒ，θは対象物までの距離とその角度を、添字０，１，２は対象物がそれぞれゴールライン、最近傍の壁、第二近傍の壁、添字３は隣のロボットを示している。ＢＲＬの出力はロボットの左右のモータ回転速度Ｏ＝｛m_0，m_1｝の２次元である。ＳＶＭの設定は、ライブラリＬＩＢＳＶＭ（http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html参照）を改良したものを使用し、各パラメータはチューニングにより規定した。

【００６６】

≪実験結果１≫

全１００試行のうち、７６試行において学習に成功した。この７６試行に関して、Ａ～

Ｈまでの各ＳＶＭによる行動選択をそれぞれ行い、７６試行中タスクを達成できた試行の成功率を図４に示す。図４のＢＲＬとはランダム行動をせず、学習成功時のみの知識をベイズ判別法により行動決定を行った場合のことである。Ｃ，Ｄ，Ｅ，Ｆの場合がタスク成功率が高く、それぞれ７８％，７１％，７８％，７１％であった。ＢＲＬのみで環境変化のタスク達成率は５１％であることから、カーネル関数がK_{poly}、K_{RBF}を用いた場合が判別性能が高いとわかる。ＢＲＬに比べ、ＳＶＭを使用したＡ～Ｈのすべての場合でタスク成功率が高いことからＳＶＭの判別性能は比較的高いとも言える。また、One-versus-AllとOne-versus-OneをK_{poly}とK_{RBF}の場合で比較すると、若干であるがOne-versus-Allの方が成功率が高い。これは一般的にOne-versus-Allはクラス数が１０００などの多い場合に有効であるとも言われるが、クラス数（ルール数）が最大１００のＢＲＬにおいては有効性に差はない。

【００６７】

≪実験結果２≫

実験１の結果から、実験２ではＣ，Ｄ，Ｅ，Ｆの４パターンの場合について、機械学習システム１を用いて学習を行った。１００試行ずつ行った結果、Ｃ，Ｄ，Ｅ，Ｆはそれぞれ７８，７９，７５，７９試行において学習に成功した。ＢＲＬのみの学習成功回数７６試行と比較し、大きな差はないように思われる。学習成功までに要したエピソード数の平均と標準偏差を図５に示す。この結果にＴ検定を行ったところ、ＢＲＬのみと比較しＤとＦの場合に関しては有意水準１％において差があることが示された。つまり、ＤもしくはＦの場合、学習するまでに要する収束速度が早いことが示された。

【００６８】

実験１と２の結果から総合的に判断して、ＤもしくはＦのパターンのＳＶＭを用いることが行動獲得に有効であり、識別精度が高いと言える。

【産業上の利用可能性】

【００６９】

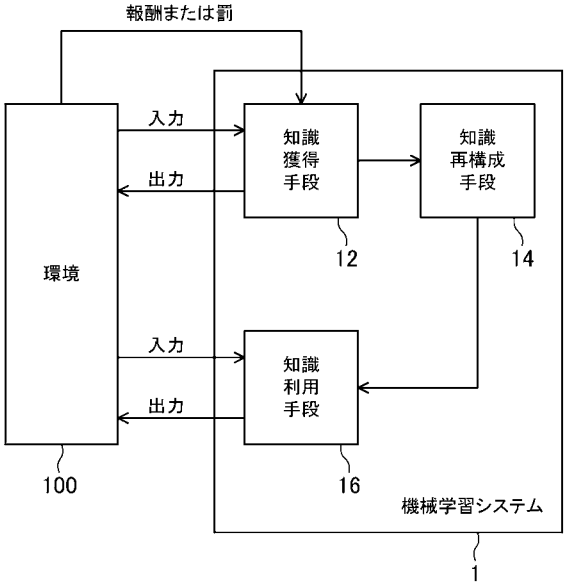
本発明に係る機械学習システムおよび機械学習方法は、頑健性に優れ、環境変化にも対応可能であるため、マルチロボットシステムなどに有用である。また、ロボットに限らず、パターン認識における強化学習にも有用である。

【符号の説明】

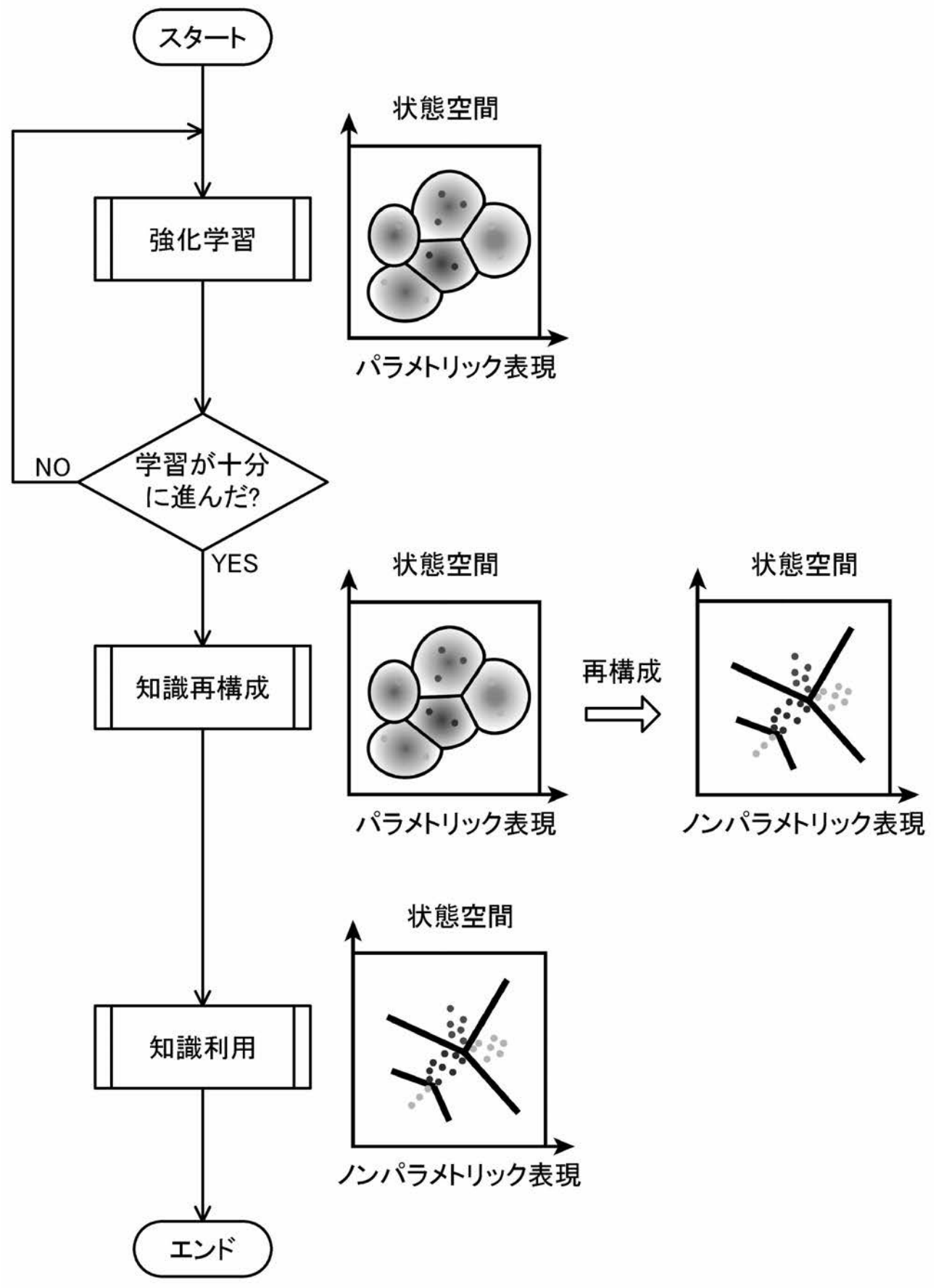
【００７０】

１　　機械学習システム
１２　　知識獲得手段
１４　　知識再構成手段
１６　　知識利用手段
１００　環境

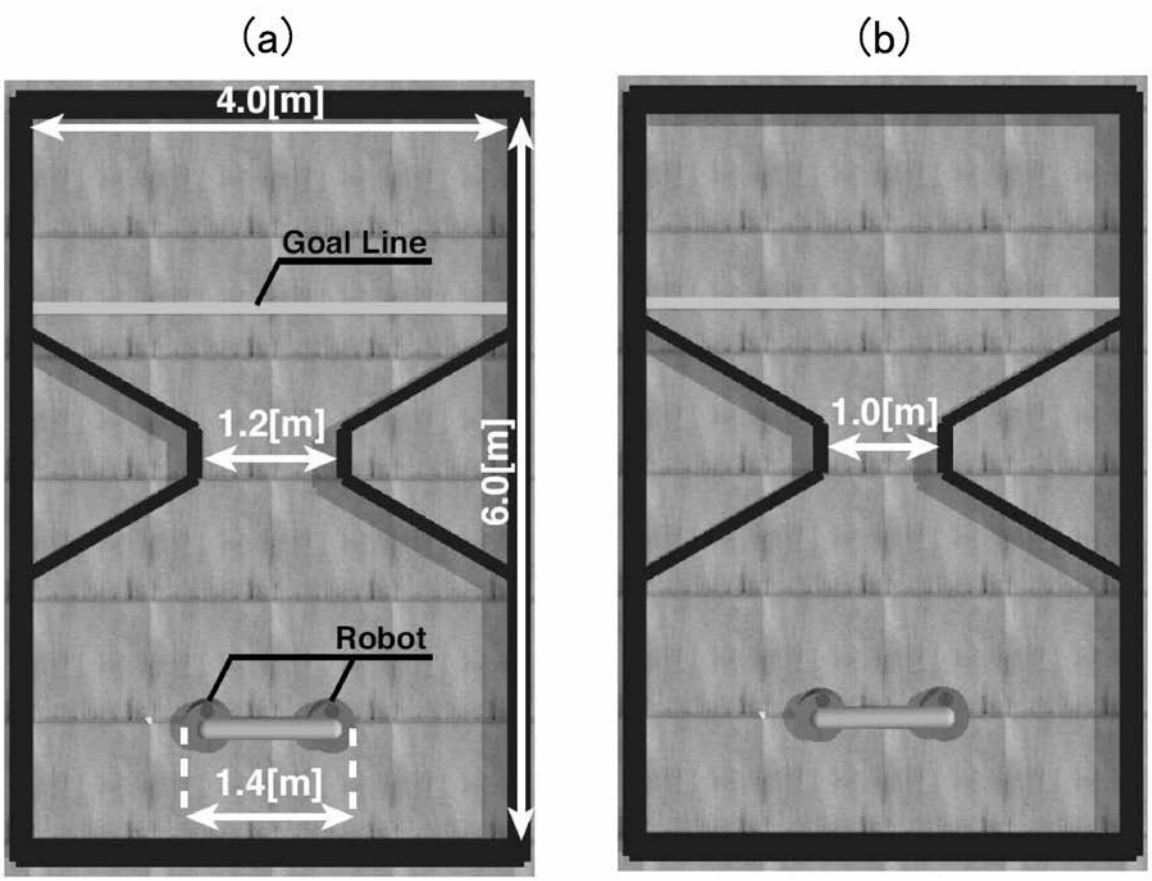
【図１】


報酬または罰
入力
出力
知識
獲得
手段
知識
再構成
手段
環境
12
14
入力
出力
知識
利用
手段
100
16
機械学習システム
1

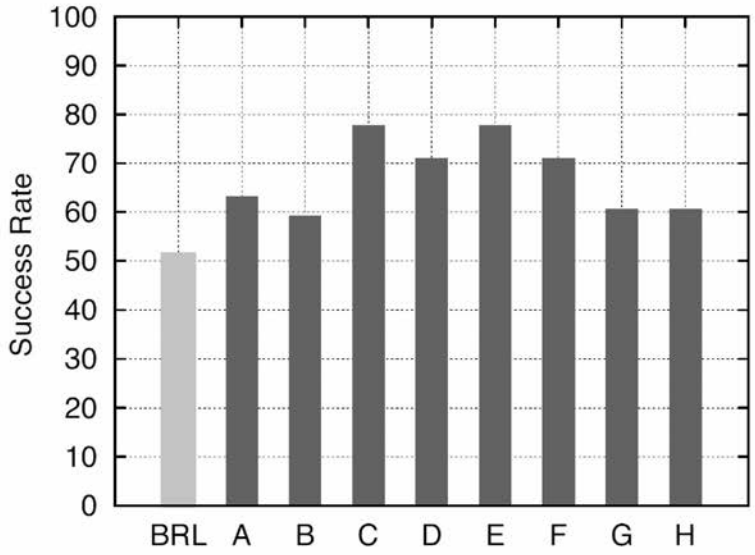
【図２】


スタート
強化学習
学習が十分に進んだ?
NO
YES
知識再構成
知識利用
エンド
状態空間
パラメトリック表現
状態空間
パラメトリック表現
再構成
状態空間
ノンパラメトリック表現
状態空間
ノンパラメトリック表現

【図３】


(a)
(b)
4.0[m]
Goal Line
1.2[m]
6.0[m]
Robot
1.4[m]
1.0[m]

【図４】


Success Rate
100
90
80
70
60
50
40
30
20
10
0
BRL
A
B
C
D
E
F
G
H

【図５】