

BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

speed, and the automatic driving robot 4, which is mounted in the vehicle 2 and causes the vehicle 2 to travel on the basis of the operation, wherein advance reinforcement learning of the operation inference learning model 70 is carried out through the application, to the operation inference learning model 70, of a simulated travel state output by a vehicle learning model 60, and after the advance reinforcement learning has finished, additional reinforcement learning of the operation inference learning model 70 is carried out through the application, to the operation inference learning model 70, of a travel state acquired by causing the vehicle 2 to travel on the basis of the operation inferred by the operation inference learning model 70.

(57) 要約: 車速を含む車両2の走行状態を基に、前記車両2を規定された指令車速に従って走行させるような、前記車両2の操作を推論する操作推論学習モデル70と、前記車両2に搭載されて、前記操作を基に当該車両2を走行させる自動操縦ロボット4を備え、前記操作推論学習モデル70を強化学習する、自動操縦ロボット4を制御する操作推論学習モデル70の学習システム10であって、前記車両学習モデル60が出力した前記模擬走行状態を前記操作推論学習モデル70に適用することで、前記操作推論学習モデル70を事前に強化学習し、当該事前の強化学習が終了した後に、前記操作推論学習モデル70が推論した前記操作を基に前記車両2を走行させて取得された前記走行状態を前記操作推論学習モデル70に適用することで、前記操作推論学習モデル70を更に強化学習する、自動操縦ロボット4を制御する操作推論学習モデル70の学習システム10を提供する。

明 細 書

発明の名称：

自動操縦ロボットを制御する操作推論学習モデルの学習システム及び学習方法

技術分野

[0001] 本発明は、自動操縦ロボットを制御する操作推論学習モデルの学習システム及び学習方法に関する。

背景技術

[0002] 一般に、普通自動車などの車両を製造、販売する際には、国や地域により規定された、特定の走行パターン（モード）により車両を走行させた際の燃費や排出ガスを測定し、これを表示する必要がある。

モードは、例えば、走行開始から経過した時間と、その時に到達すべき車速との関係として、グラフにより表わすことが可能である。この到達すべき車速は、車両へ与えられる達成すべき速度に関する指令という観点で、指令車速と呼ばれることがある。

上記のような、燃費や排出ガスに関する試験は、シャシーダイナモメータ上に車両を載置し、車両に搭載された自動操縦ロボット、所謂ドライプロボット（登録商標）により、モードに従って車両を運転させることにより行われる。

[0003] 指令車速には、許容誤差範囲が規定されている。車速が許容誤差範囲を逸脱すると、その試験は無効となるため、自動操縦ロボットの制御には、指令車速への高い追従性が求められる。このため、自動操縦ロボットを、例えば強化学習により学習された学習モデルを用いて制御することがある。

例えば、特許文献1には、人間らしいペダル操作を行うドライバモデルを強化学習によって構築することが可能な車輛用走行シミュレーション装置、ドライバモデル構築方法及びドライバモデル構築プログラムが開示されている。

より詳細には、車両用走行シミュレーション装置は、ドライバモデルのゲインの値を変更させながら、車両モデルを複数回走行させ、この時に変更されたゲインの値を報酬値に基づいて評価することによって、ドライバモデルのゲインの設定を自動的に行う。上記ゲインの値は、車速の追従性を評価する車速報酬関数のみならず、アクセルペダルの操作の滑らかさを評価するアクセル報酬関数、ブレーキペダルの操作の滑らかさを評価するブレーキ報酬関数によっても評価が行われる。

特許文献1等において用いられる車両モデルとしては、通常、車両の各構成要素に対して、動作を模した物理モデルを各々作成し、これらを組み合わせた物理モデルとして作成される。

先行技術文献

特許文献

[0004] 特許文献1：特開2014-115168号公報

発明の概要

発明が解決しようとする課題

[0005] 特許文献1に開示されたような装置においては、車両の操作を推論する操作推論学習モデルを、車両モデルを基に学習している。このため、車両モデルの再現精度が低いと、操作推論学習モデルをどれだけ精密に学習させたとしても、操作推論学習モデルが推論する操作が、実際の車両にそぐわないものとなり得る。特に、物理モデルの作成は、実車両の微細なパラメータを解析してこれを反映させる必要があるために、これを用いて精度が高い車両モデルを構築するのは容易ではない。このため、特に車両モデルとして物理モデルを使用すると、操作推論学習モデルにより出力する操作の精度を高めるのが難しい。

[0006] 他方、操作推論学習モデルの強化学習時に、車両モデルに替えて、実際の車両を使用することが考えられる。すなわち、操作推論学習モデルによる操作の推論、当該操作による実車両の操作、操作の結果である走行実績として

の実車両の走行状態の蓄積、及び蓄積された走行状態を用いた操作推論学習モデルの更なる学習を、操作推論学習モデルの操作推論の精度が高まるまで繰り返すことで、操作推論学習モデルを強化学習する。この場合においては、最終的に生成される操作推論学習モデルを、実際の車両の試験に十分に適用可能な程度の精度にすることができる。

しかし、強化学習においては、上記のように学習モデルの学習と、学習途中の学習モデルが推論した操作を用いた走行状態の取得を繰り返して、学習モデルの学習を進める。したがって、学習モデルは、学習の初期段階においては、例えばペダルを極端に高い頻度で操作するような、人間には不可能で、実車両に負担がかかる、好ましくない操作を出力する可能性がある。

[0007] 本発明が解決しようとする課題は、強化学習中の操作推論学習モデルによる好ましくない車両操作の出力を低減して実車両の負担を低減可能で、かつ操作推論学習モデルにより出力する操作の精度を向上可能な、自動操縦ロボット（ドライプロボット）を制御する操作推論学習モデルの学習システム及び学習方法を提供することである。

課題を解決するための手段

[0008] 本発明は、上記課題を解決するため、以下の手段を採用する。すなわち、本発明は、車速を含む車両の走行状態を基に、前記車両を規定された指令車速に従って走行させるような、前記車両の操作を推論する操作推論学習モデルと、前記車両に搭載されて、前記操作を基に当該車両を走行させる自動操縦ロボットを備え、前記操作推論学習モデルを強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習システムであって、前記車両の実際の走行実績を基に前記車両を模擬動作するように機械学習され、前記操作推論学習モデルが推論した前記操作を基に、前記車両を模した前記走行状態である模擬走行状態を出力する、車両学習モデルを備え、前記車両学習モデルが出力した前記模擬走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを事前に強化学習し、当該事前の強化学習が終了した後に、前記操作推論学習モデルが推論した前記操作を基に前記車両を走

行させて取得された前記走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを更に強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習システムを提供する。

[0009] また、本発明は、車速を含む車両の走行状態を基に、前記車両を規定された指令車速に従って走行させるような、前記車両の操作を推論する操作推論学習モデルと、前記車両に搭載されて、前記操作を基に当該車両を走行させる自動操縦ロボットに関し、前記操作推論学習モデルを強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習方法であって、前記車両の実際の走行実績を基に前記車両を模擬動作するように機械学習された車両学習モデルにより、前記操作推論学習モデルが推論した前記操作を基に、前記車両を模した前記走行状態である模擬走行状態を出力し、当該模擬走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを事前に強化学習し、前記事前の強化学習が終了した後に、前記操作推論学習モデルが推論した前記操作を基に前記車両を走行させて取得された前記走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを更に強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習方法を提供する。

発明の効果

[0010] 本発明によれば、強化学習中の操作推論学習モデルによる好ましくない車両操作の出力を低減して実車両の負担を低減可能で、かつ操作推論学習モデルにより出力する操作の精度を向上可能な、自動操縦ロボット（ドライプロボット）を制御する操作推論学習モデルの学習システム及び学習方法を提供することができる。

図面の簡単な説明

[0011] [図1]本発明の実施形態における、自動操縦ロボット（ドライプロボット）を用いた試験環境の説明図である。

[図2]上記実施形態における自動操縦ロボットを制御する操作推論学習モデルの学習システムの、車両学習モデルの学習時における処理の流れを記したブ

ロック図である。

[図3]上記車両学習モデルのブロック図である。

[図4]上記自動操縦ロボットを制御する操作推論学習モデルの学習システムの、操作推論学習モデルの事前学習時における処理の流れを記したブロック図である。

[図5]上記操作推論学習モデルのブロック図である。

[図6]上記操作推論学習モデルの強化学習に用いられる、価値推論学習モデルのブロック図である。

[図7]上記自動操縦ロボットを制御する操作推論学習モデルの学習システムの、操作推論学習モデルの事前学習が終了した後の強化学習時における処理の流れを記したブロック図である。

[図8]上記実施形態における自動操縦ロボットを制御する操作推論学習モデルの学習方法のフローチャートである。

発明を実施するための形態

[0012] 以下、本発明の実施形態について図面を参照して詳細に説明する。

本実施形態においては、自動操縦ロボットとしては、ドライプロボット（登録商標）を用いているため、以下、自動操縦ロボットをドライプロボットと記載する。

[0013] 図1は、実施形態におけるドライプロボットを用いた試験環境の説明図である。試験装置1は、車両2、シャシーダイナモメータ3、及びドライプロボット4を備えている。

車両2は、床面上に設けられている。シャシーダイナモメータ3は、床面の下方に設けられている。車両2は、車両2の駆動輪2aがシャシーダイナモメータ3の上に載置されるように、位置づけられている。車両2が走行し駆動輪2aが回転する際には、シャシーダイナモメータ3が反対の方向に回転する。

ドライプロボット4は、車両2の運転席2bに搭載されて、車両2を走行させる。ドライプロボット4は、第1アクチュエータ4cと第2アクチュエ

ータ 4 d を備えており、これらはそれぞれ、車両 2 のアクセルペダル 2 c とブレーキペダル 2 d に当接するように設けられている。

[0014] ドライブロボット 4 は、後に詳説する学習制御装置 1 1 によって制御されている。学習制御装置 1 1 は、ドライブロボット 4 の第 1 アクチュエータ 4 c と第 2 アクチュエータ 4 d を制御することにより、車両 2 のアクセルペダル 2 c とブレーキペダル 2 d の開度を変更、調整する。

学習制御装置 1 1 は、ドライブロボット 4 を、車両 2 が規定された指令車速に従って走行するように制御する。すなわち、学習制御装置 1 1 は、車両 2 のアクセルペダル 2 c とブレーキペダル 2 d の開度を変更することで、規定された走行パターン（モード）に従うように、車両 2 を走行制御する。より詳細には、学習制御装置 1 1 は、走行開始から時間が経過するに従い、各時間に到達すべき車速である指令車速に従うように、車両 2 を走行制御する。

[0015] 学習制御システム（学習システム）1 0 は、上記のような試験装置 1 と学習制御装置 1 1 を備えている。

学習制御装置 1 1 は、ドライブロボット制御部 2 0 と学習部 3 0 を備えている。

ドライブロボット制御部 2 0 は、ドライブロボット 4 の制御を行うための制御信号を生成し、ドライブロボット 4 に送信することで、ドライブロボット 4 を制御する。学習部 3 0 は、後に説明するような機械学習を行い、車両学習モデル、操作推論学習モデル、及び価値推論学習モデルを生成する。上記のような、ドライブロボット 4 の制御を行うための制御信号は、操作推論学習モデルにより生成される。

ドライブロボット制御部 2 0 は、例えば、ドライブロボット 4 の筐体外部に設けられた、コントローラ等の情報処理装置である。学習部 3 0 は、例えばパーソナルコンピュータ等の情報処理装置である。

[0016] 図 2 は、学習制御システム 1 0 のブロック図である。図 2 においては、各構成要素を結ぶ線は、上記車両学習モデルを機械学習する際にデータの送受

信があるもののみが示されており、したがって構成要素間の全てのデータの送受信を示すものではない。

試験装置 1 は、既に説明したような車両 2、シャシーダイナモメータ 3、及びドライブロボット 4 に加え、車両状態計測部 5 を備えている。車両状態計測部 5 は、車両 2 の状態を計測する各種の計測装置である。車両状態計測部 5 としては、例えばアクセルペダル 2 c やブレーキペダル 2 d の操作量を計測するためのカメラや赤外線センサなどであり得る。

本実施形態においては、ドライブロボット 4 が第 1 及び第 2 アクチュエータ 4 c、4 d を制御することでこれらペダル 2 c、2 d を操作する。したがって、車両状態計測部 5 に依らずとも、例えば第 1 及び第 2 アクチュエータ 4 c、4 d の制御量等を基に、ペダル 2 c、2 d の操作量を把握することができる。このため、本実施形態においては、車両状態計測部 5 は必須の構成ではない。しかし、後に変形例として記載するような、例えばドライブロボット 4 に替えて人間が車両 2 を操縦する際にペダル 2 c、2 d の操作量を把握したい場合や、エンジン回転数やギア状態、エンジン温度等の車両 2 の状態を直接計測して把握したい場合には必要となる。

[0017] ドライブロボット制御部 20 は、ペダル操作パターン生成部 21、車両操作制御部 22、及び駆動状態取得部 23 を備えている。学習部 30 は、指令車速生成部 31、推論データ成形部 32、学習データ成形部 33、学習データ生成部 34、学習データ記憶部 35、強化学習部 40、及び試験装置モデル 50 を備えている。強化学習部 40 は、操作内容推論部 41、状態行動価値推論部 42、及び報酬計算部 43 を備えている。試験装置モデル 50 は、ドライブロボットモデル 51、車両モデル 52、及びシャシーダイナモメータモデル 53 を備えている。

学習制御装置 11 の、学習データ記憶部 35 以外の各構成要素は、例えば上記の各情報処理装置内の CPU により実行されるソフトウェア、プログラムであってよい。また、学習データ記憶部 35 は、上記各情報処理装置内外に設けられた半導体メモリや磁気ディスクなどの記憶装置により実現されて

いてよい。

[0018] 後に説明するように、操作内容推論部 4 1 は、ある時刻における走行状態を基に、指令車速に従うような、当該時刻よりも後の車両 2 の操作を推論する。この、車両 2 の操作の推論を効果的に行うために、特に操作内容推論部 4 1 は、後に説明するように機械学習器を備えており、推論した操作に基づいたドライプロボット 4 の操作の後の時刻における走行状態に基づいて計算された報酬を基に機械学習器を強化学習して学習モデル（操作推論学習モデル）7 0 を生成する。操作内容推論部 4 1 は、性能測定のために実際に車両 2 を走行制御させる際には、この学習が完了した操作推論学習モデル 7 0 を使用して、車両 2 の操作を推論する。

すなわち、学習制御システム 1 0 は大別して、強化学習時における操作の学習と、性能測定のために車両を走行制御させる際における操作の推論の、2 通りの動作を行う。説明を簡単にするために、以下ではまず、操作の学習時における、学習制御システム 1 0 の各構成要素の説明をした後に、車両の性能測定に際して操作を推論する場合での各構成要素の挙動について説明する。

[0019] まず、操作の学習時における、学習制御装置 1 1 の構成要素の挙動を説明する。

学習制御装置 1 1 は、操作の学習に先立ち、学習時に使用する走行実績データ（走行実績）を、走行実績として収集する。詳細には、ドライプロボット制御部 2 0 が、アクセルペダル 2 c 及びブレーキペダル 2 d の、車両特性計測用の操作パターンを生成して、これにより車両 2 を走行制御し、走行実績データを収集する。

ペダル操作パターン生成部 2 1 は、ペダル 2 c、2 d の、車両特性計測用の操作パターンを生成する。ペダル操作パターンとしては、例えば車両 2 と類似する他の車両において、W L T C (W o r l d w i d e h a r m o n i z e d L i g h t v e h i c l e s T e s t C y c l e) モードなどによって走行した際のペダル操作の実績値を使用することができる。

ペダル操作パターン生成部 2 1 は、生成したペダル操作パターンを、車両操作制御部 2 2 へ送信する。

[0020] 車両操作制御部 2 2 は、ペダル操作パターン生成部 2 1 から、ペダル操作パターンを受信し、これを、ドライブロボット 4 の第 1 及び第 2 アクチュエータ 4 c、4 d への指令に変換して、ドライブロボット 4 に送信する。

ドライブロボット 4 は、アクチュエータ 4 c、4 d への指令を受信すると、これに基づいて車両 2 をシャシーダイナモメータ 3 上で走行させる。

駆動状態取得部 2 3 は、例えばアクチュエータ 4 c、4 d の位置等の、ドライブロボット 4 の実際の駆動状態を取得する。車両 2 が走行することにより、車両 2 の走行状態は逐次変化する。駆動状態取得部 2 3 と、車両状態計測部 5、及びシャシーダイナモメータ 3 に設けられた様々な計測器により、車両 2 の走行状態が計測される。例えば、駆動状態取得部 2 3 は上記のように、アクセルペダル 2 c の検出量と、ブレーキペダル 2 d の検出量を、走行状態として計測する。また、シャシーダイナモメータ 3 に設けられた計測器は、車速を走行状態として計測する。

計測された車両 2 の走行状態は、学習部 3 0 の学習データ成形部 3 3 へ送信される。

学習データ成形部 3 3 は、車両 2 の走行状態を受信し、受信したデータを後の様々な学習において使用されるフォーマットに変換して、走行実績データとして学習データ記憶部 3 5 に保存する。

[0021] 車両 2 の走行状態すなわち走行実績データの収集が終了すると、学習データ生成部 3 4 は学習データ記憶部 3 5 から走行実績データを取得し、適切なフォーマットに成形して、試験装置モデル 5 0 に送信する。

試験装置モデル 5 0 の車両モデル 5 2 は、学習データ生成部 3 4 から成形された走行実績データを取得し、これを用いて機械学習器 6 0 を機械学習して、車両学習モデル 6 0 を生成する。車両学習モデル 6 0 は、車両 2 の実際の走行実績である走行実績データを基に車両 2 を模擬動作するように機械学習され、車両 2 に対する操作を受信すると、これを基に、車両 2 を模した走

行状態である模擬走行状態を出力する。すなわち、車両モデル52の機械学習器60は、人工知能ソフトウェアの一部であるプログラムモジュールとして利用される、適切な学習パラメータが学習された学習済みモデル60を生成するものである。

本実施形態においては、車両学習モデル60は、ニューラルネットワークで実現され、所定の時刻を基点とした走行状態が学習データとして、及び所定の時刻より後の時刻の走行実績が教師データとして入力され、後の時刻の模擬走行状態を出力してこれを教師データと比較することで機械学習される。

以下、説明を簡単にするため、車両モデル52が備えている機械学習器と、これが学習されて生成される学習モデルとともに、車両学習モデル60と呼称する。

[0022] 図3は、車両学習モデル60のブロック図である。本実施形態においては、車両学習モデル60は、中間層を3層とした全5層の全結合型のニューラルネットワークにより実現されている。車両学習モデル60は、入力層61、中間層62、及び出力層63を備えている。図3においては、各層が矩形として描かれており、各層に含まれるノードは省略されている。

[0023] 本実施形態においては、車両学習モデル60の入力となる走行状態は、任意の所定の時刻を基点として、所定の第1時間だけ過去から基点となる時刻までの間の、車速の系列を含む。また、本実施形態においては、車両学習モデル60の入力となる走行状態は、基点となる時刻から所定の第2時間だけ将来の時刻までの間の、アクセルペダル2cの操作量の系列、及びブレーキペダル2dの操作量の系列を含む。

入力層61は、上記のような車速の系列である車速系列*i*1、アクセルペダル2cの操作量の系列であるアクセルペダル系列*i*2、及びブレーキペダル2dの操作量の系列であるブレーキペダル系列*i*3の各々に対応する入力ノードを備えている。

上記のように、各入力*i*1、*i*2、*i*3は系列であり、それぞれ、複数の

値により実現されている。例えば、図3においては、一つの矩形として示されている、車速系列 i 1 に対応する入力は、実際には、車速系列 i 1 の複数の値の各々に対応するように、入力ノードが設けられている。

車両モデル52は、各入力ノードに、対応する走行実績データの値を格納する。

[0024] 中間層62は、第1中間層62a、第2中間層62b、及び第3中間層62cを備えている。

中間層62の各ノードにおいては、前段の層（例えば、第1中間層62aの場合は入力層61、第2中間層62bの場合は第1中間層62a）の各ノードから、この前段の層の各ノードに格納された値と、前段の層の各ノードから当該中間層62のノードへの重みを基にした演算がなされて、当該中間層62のノード内に演算結果が格納される。

出力層63においても、中間層62の各々と同様な演算が行われ、出力層63に備えられた各出力ノードに演算結果が格納される。

本実施形態においては、車両学習モデル60の出力は、基点となる時刻から所定の第3時間だけ将来の時刻までの間の、推定された車速の系列である。この推定車速系列 o は、系列であるため、複数の値により実現されている。例えば、図3においては、一つの矩形として示されている、推定車速系列 o に対応する出力は、実際には、推定車速系列 o の複数の値の各々に対応するように、出力ノードが設けられている。

[0025] 車両学習モデル60においては、上記のように所定の時刻の走行実績が走行状態 i 1、 i 2、 i 3として入力されて、後の時刻の適切な推定車速系列 o を、車両2の走行を模した走行状態である模擬走行状態 o として出力することができるように学習がなされる。

より詳細には、車両モデル52は、別途学習データ記憶部35から学習データ生成部34を介して送信された、基点となる所定の時刻から所定の第3時間だけ将来の時刻までの間の走行実績、本実施形態においては車速系列の正解値を、教師データとして受信する。車両モデル52は、教師データと、

車両学習モデル60が出力した推定車速系列oの平均二乗誤差が小さくなるように、重みやバイアスの値等、ニューラルネットワークを構成する各パラメータの値を、誤差逆伝搬法、確率的勾配降下法により調整する。

車両モデル52は、車両学習モデル60の学習を繰り返しつつ、教師データと推定車速系列oの最小二乗誤差を都度計算し、これが所定の値よりも小さければ、車両学習モデル60の学習を終了する。

[0026] 車両学習モデル60の学習が終了すると、学習制御システム10の強化学習部40は、操作内容推論部41に設けられた、車両2の操作を推論する操作推論学習モデル70を事前学習する。図4は、事前学習時のデータの送受信関係が示された学習制御システム10のブロック図である。操作推論学習モデル70は、機械学習器が学習されることにより、人工知能ソフトウェアの一部であるプログラムモジュールとして利用される、適切な学習パラメータが学習された学習済みモデルとなる。

学習制御システム10は、既に学習が終了した車両学習モデル60が出力した模擬走行状態を操作推論学習モデル70に適用することで、操作推論学習モデル70を事前に強化学習する。後に説明するように、操作推論学習モデル70の強化学習が進行して事前の強化学習が終了した後に、操作推論学習モデル70の出力した操作を基に実際に車両2を走行させて取得された走行状態を操作推論学習モデル70に適用することで、操作推論学習モデル70を更に強化学習する。このように、学習制御システム10は、操作推論学習モデル70の学習段階に応じて、推論した操作の実行対象及び走行状態の取得対象を、車両学習モデル60から実車両2へと変更する。

[0027] 後に説明するように、操作内容推論部41は、現時点から所定の第3時間だけ将来の時刻までの間の車両2の操作を出力し、これをドライバロボットモデル51に送信する。本実施形態において、操作内容推論部41は、特にアクセルペダル2c及びブレーキペダル2dの操作の系列を出力する。

車両学習モデル60の学習により、試験装置モデル50は、全体として試験装置1の各々を模擬動作するように構成されている。試験装置モデル50

は、操作の系列を受信する。

[0028] ドライブロボットモデル51は、ドライブロボット4を模擬動作するように構成されている。ドライブロボットモデル51は、受信した操作を基に、学習が終了した車両学習モデル60の入力となるアクセルペダル系列i2とブレーキペダル系列i3を生成し、車両モデル52に送信する。

シャシーダイナモメータモデル53は、シャシーダイナモメータ3を模擬動作するように構成されている。シャシーダイナモメータ3は、模擬走行中の車両学習モデル60の車速を検出しつつ、これを内部に随時記録している。シャシーダイナモメータモデル53は、この過去の車速の記録から車速系列i1を生成し、車両モデル52に送信する。

車両モデル52は、車速系列i1、アクセルペダル系列i2、及びブレーキペダル系列i3を受信してこれらを車両学習モデル60に入力する。車両学習モデル60が推定車速系列oを出力すると、車両モデル52は推定車速系列oを推論データ成形部32に送信する。

シャシーダイナモメータモデル53は、車両学習モデル60からこの時点における車速を検出して車速系列i1を更新し、推論データ成形部32に送信する。

[0029] 指令車速生成部31は、モードに関する情報に基づいて生成された、指令車速を保持している。指令車速生成部31は、現時点から所定の第4時間だけ将来の時刻までの間に、車両学習モデル60が従うべき指令車速の系列を生成し、推論データ成形部32に送信する。

推論データ成形部32は、推定車速系列oと指令車速系列を受信し、適切に成形した後に強化学習部40に送信する。

[0030] 強化学習部40は、過去に送信したアクセルペダル2cとブレーキペダル2dの操作を保持している。強化学習部40は、これらの送信した操作を、車両学習モデル60が実際にこれに従った結果としての検出値として見做し、これらアクセルペダル2cとブレーキペダル2dの操作の系列を基に、過去のアクセルペダル検出量の系列とブレーキペダル検出量の系列を生成する

。強化学習部40は、これらを推定車速系列oと指令車速系列とともに、走行状態として操作内容推論部41に送信する。

[0031] 操作内容推論部41は、ある時刻において走行状態を受信すると、これを基に、学習中の操作推論学習モデル70により、当該時刻より後の操作の系列を推論する。図5は、操作推論学習モデル70のブロック図である。

操作推論学習モデル70の入力層71においては、入力ノードが、例えばアクセルペダル検出量s1、ブレーキペダル検出量s2から、指令車速sNに至るまでの、走行状態sの各々に対応するように設けられている。操作推論学習モデル70は、車両学習モデル60と同様な構造のニューラルネットワークにより実現されているため、構造上の詳細な説明を割愛する。

操作推論学習モデル70の出力層73においては、出力ノードの各々は、操作aの各々に対応するように設けられている。本実施形態においては、操作の対象は、アクセルペダル2cとブレーキペダル2dであり、これに対応して、操作aは、例えばアクセルペダル操作の系列a1とブレーキペダル操作の系列a2となっている。

[0032] 操作内容推論部41は、このようにして生成されたアクセルペダル操作a1とブレーキペダル操作a2をドライブロボットモデル51に送信する。ドライブロボットモデル51は、これを基にアクセルペダル系列i2とブレーキペダル系列i3を生成して車両学習モデル60に送信し、車両学習モデル60は次の車速を推論する。この、次の車速を基にして、次の走行状態sが生成される。

操作推論学習モデル70の学習、すなわち誤差逆伝搬法、確率的勾配降下法によるニューラルネットワークを構成する各パラメータの値の調整は、現段階においては行われず、操作推論学習モデル70は操作aを推論するのみである。操作推論学習モデル70の学習は、後に、価値推論学習モデル80の学習に伴って行われる。

[0033] 報酬計算部43は、走行状態sと、これに対応して操作推論学習モデル70により推論された操作a、及び当該操作aを基に新たに生成された走行状

態 s を基に、適切に設計された式により報酬を計算する。報酬は、操作 a 、及びこれに伴う新たに生成された走行状態 s が望ましくないほど小さい値を、望ましいほど大きい値を、有するように設計されている。後述する状態行動価値推論部 42 は、行動価値を、報酬が大きいほどこれが高くするように計算し、操作推論学習モデル 70 はこの行動価値が高くなるような操作 a を出力するように、強化学習が行われる。

報酬計算部 43 は、走行状態 s 、これに対応して推論された操作 a 、当該操作 a を基に新たに生成された走行状態 s 、及び計算した報酬を、学習データ成形部 33 に送信する。学習データ成形部 33 は、これらを適切に成形して学習データ記憶部 35 に保存する。これらのデータは、後述する価値推論学習モデル 80 の学習に使用される。

このようにして、操作内容推論部 41 による操作 a の推論と、この操作 a に対応した、車両モデル 52 による推定車速系列 o の推論、及び報酬の計算が、価値推論学習モデル 80 の学習に十分なデータが蓄積されるまで、繰り返し行われる。

[0034] 学習データ記憶部 35 に、価値推論学習モデル 80 の学習に十分な量の走行データが蓄積されると、状態行動価値推論部 42 は価値推論学習モデル 80 を学習する。価値推論学習モデル 80 は、機械学習器が学習されることにより、人工知能ソフトウェアの一部であるプログラムモジュールとして利用される、適切な学習パラメータが学習された学習済みモデルとなる。

強化学習部 40 は全体として、操作推論学習モデル 70 が推論した操作 a がどの程度適切であったかを示す行動価値を計算し、操作推論学習モデル 70 が、この行動価値が高くなるような操作 a を出力するように、強化学習を行う。行動価値は、走行状態 s と、これに対する操作 a を引数として、報酬が大きいほど行動価値 Q を高くするように設計された関数 Q として表わされる。本実施形態においては、この関数 Q の計算を、走行状態 s と操作 a を入力として、行動価値 Q を出力するように設計された、関数近似器としての学習モデル 80 により行う。

[0035] 状態行動価値推論部42は、学習データ記憶部35から、学習データ生成部34を介して成形された走行状態 s と操作 a を受信し、価値推論学習モデル80を機械学習させる。図6は、価値推論学習モデル80のブロック図である。

価値推論学習モデル80の入力層81においては、入力ノードが、例えばアクセルペダル検出量 s_1 、ブレーキペダル検出量 s_2 から、指令車速 s_N に至るまでの、走行状態 s の各々と、及び、例えばアクセルペダル操作 a_1 とブレーキペダル操作 a_2 の、操作 a の各々に対応するように設けられている。価値推論学習モデル80は、車両学習モデル60と同様な構造のニューラルネットワークにより実現されているため、構造上の詳細な説明を割愛する。

価値推論学習モデル80の出力層83においては、出力ノードは、例えば1つであり、これが、計算された行動価値 Q の値に相当する。

[0036] 報酬計算部43は、TD (Temporal Difference) 誤差、すなわち、操作 a を実行する前の行動価値と、操作 a を実行した後の行動価値の誤差を小さくして、行動価値 Q として適切な値が出力されるように、重みやバイアスの値等、ニューラルネットワークを構成する各パラメータの値を、誤差逆伝搬法、確率的勾配降下法により調整する。このように、現状の操作推論学習モデル70によって推論された操作 a を適切に評価できるように、価値推論学習モデル80を学習させる。

価値推論学習モデル80の学習が終了すると、価値推論学習モデル80は、より適切な行動価値 Q の値を出力するようになる。すなわち、価値推論学習モデル80が出力する行動価値 Q の値が学習前とは変わるため、これに伴い、行動価値 Q が高くなるような操作 a を出力するように設計された操作推論学習モデル70を更新する必要がある。このため、操作内容推論部41は操作推論学習モデル70を学習する。

具体的には、状態行動価値推論部42は、例えば行動価値 Q の負値を損失関数とし、これをできるだけ小さくするような、すなわち行動価値 Q が大き

くなるような操作 a を出力するように、重みやバイアスの値等、ニューラルネットワークを構成する各パラメータの値を、誤差逆伝搬法、確率的勾配降下法により調整して、操作推論学習モデル 70 を学習させる。

操作推論学習モデル 70 が学習され更新されると、出力される操作 a が変化するため、再度走行データを蓄積し、これを基に価値推論学習モデル 80 を学習する。

このように、学習部 30 は、操作推論学習モデル 70 と価値推論学習モデル 80 の学習を繰り返すことにより、これら学習モデル 70、80 を強化学習する。

[0037] 学習部 30 は、この事前学習としての、車両学習モデル 60 を操作 a の実行対象として用いた強化学習を、所定の事前学習終了基準を満たすまで実行する。

例えば、学習部 30 は、車両学習モデル 60 を操作 a の実行対象として用いた制御で十分な走行性能が得られるまで、事前学習を実行する。例えば、学習制御システム 10 がモード走行を前提としている場合においては、車両学習モデル 60 によるモード走行において、車速指令と推定車速系列 o との誤差が所定の閾値以下となり十分に小さい値となるまで、事前学習が実行される。

あるいは、所定の時間範囲におけるアクセルペダル 2 c やブレーキペダル 2 d の操作回数、操作量や変化率が、所定の閾値以下になった場合は、実際に車両 2 で試験を実行しても車両 2 に大きな負荷を与える可能性が低いと判断して、事前学習を終了するようにしてもよい。

[0038] 操作推論学習モデル 70 及び価値推論学習モデル 80 の、車両学習モデル 60 を操作 a の実行対象として用いた事前学習が終了すると、学習部 30 は、車両学習モデル 60 に替えて、実車両 2 を操作 a の実行対象として、操作推論学習モデル 70 及び価値推論学習モデル 80 を更に強化学習する。図 7 は、事前学習が終了した後の強化学習時におけるデータの送受信関係が示された学習制御システム 10 のブロック図である。

[0039] 操作内容推論部 4 1 は、現時点から所定の第 3 時間だけ将来の時刻までの間の車両 2 の操作 a を出力し、これを車両操作制御部 2 2 に送信する。

車両操作制御部 2 2 は、受信した操作 a を、ドライブロボット 4 の第 1 及び第 2 アクチュエータ 4 c、4 d への指令に変換して、ドライブロボット 4 に送信する。

ドライブロボット 4 は、アクチュエータ 4 c、4 d への指令を受信すると、これに基づいて車両 2 をシャシーダイナモメータ 3 上で走行させる。

シャシーダイナモメータ 3 は、車両 2 の車速を検出し、車速系列を生成して推論データ成形部 3 2 に送信する。

指令車速生成部 3 1 は、指令車速系列を生成して推論データ成形部 3 2 に送信する。

推論データ成形部 3 2 は、車速系列と指令車速系列を受信し、適切に成形した後に走行状態として、強化学習部 4 0 に送信する。

[0040] 強化学習部 4 0 は、車両モデル 5 2 により生成される推定車速系列 o の替わりに上記の車速系列を用いて、図 4 を用いて説明した事前学習時と同様に、上記のように実車両 2 を操作 a の実行対象として用いて学習データを学習データ記憶部 3 5 に蓄積する。強化学習部 4 0 は、十分な量の走行データが蓄積されると、価値推論学習モデル 8 0 を学習し、その後操作推論学習モデル 7 0 を学習する。

学習部 3 0 は、学習データの蓄積と、操作推論学習モデル 7 0 と価値推論学習モデル 8 0 の学習を繰り返すことにより、これら学習モデル 7 0、8 0 を強化学習する。

[0041] 学習部 3 0 は、車両 2 を操作 a の実行対象として用いた強化学習を、所定の学習終了基準を満たすまで実行する。

例えば、学習部 3 0 は、車両 2 を操作 a の実行対象として用いた制御で十分な走行性能が得られるまで、事前学習を実行する。例えば、学習制御システム 1 0 がモード走行を前提としている場合においては、車両 2 によるモード走行において、車速指令と実際にシャシーダイナモメータ 3 により検出さ

れる車速との誤差が所定の閾値以下となり十分に小さい値となるまで、事前学習が実行される。

[0042] 次に、車両 2 の性能測定に際して操作 a を推論する場合での、すなわち、操作推論学習モデル 70 の強化学習が終了した後における、学習制御システム 10 の各構成要素の挙動について説明する。

[0043] 駆動状態取得部 23 と、車両状態計測部 5、及びシャシーダイナモメータ 3 に設けられた様々な計測器により、車両 2 の車速、アクセルペダル 2c の検出量、ブレーキペダル 2d の検出量等が計測される。これらの値は、推論データ成形部 32 に送信される。

指令車速生成部 31 は、指令車速系列を生成して推論データ成形部 32 に送信する。

推論データ成形部 32 は、車速、アクセルペダル 2c の検出量、ブレーキペダル 2d の検出量等と、指令車速系列を受信し、適切に成形した後に走行状態として、強化学習部 40 に送信する。

操作内容推論部 41 は、走行状態を受信すると、これを基に、学習済みの操作推論学習モデル 70 により、車両 2 の操作 a を推論する。

操作内容推論部 41 は、推論した操作 a を、車両操作制御部 22 へ送信する。

車両操作制御部 22 は、操作内容推論部 41 から操作 a を受信し、この操作 a に基づき、ドライプロボット 4 を操作する。

[0044] 次に、図 1 ～図 7、及び図 8 を用いて、上記の学習制御システム 10 を用いた、ドライプロボット 4 を制御する操作推論学習モデル 70 の学習方法を説明する。図 8 は、学習方法のフローチャートである。

学習制御装置 11 は、操作の学習に先立ち、学習時に使用する走行実績データ（走行実績）を、走行実績として収集する。詳細には、ドライプロボット制御部 20 が、アクセルペダル 2c 及びブレーキペダル 2d の、車両特性計測用の操作パターンを生成して、これにより車両 2 を走行制御し、走行実績データを収集する（ステップ S1）。

車両モデル52は、学習データ生成部34から成形された走行実績データを取得し、これを用いて機械学習器60を機械学習して、車両学習モデル60を生成する（ステップS3）。

[0045] 車両学習モデル60の学習が終了すると、学習制御システム10の強化学習部40は、車両2の操作を推論する操作推論学習モデル70を事前学習する（ステップS5）。より詳細には、学習制御システム10は、既に学習が終了した車両学習モデル60が出力した模擬走行状態を操作推論学習モデル70に適用することで、操作推論学習モデル70を事前に強化学習する。

学習部30は、この事前学習としての、車両学習モデル60を操作aの実行対象として用いた強化学習を、所定の事前学習終了基準を満たすまで実行する。事前学習終了基準を満たさなければ（ステップS7のNo）、事前学習を継続する。事前学習終了基準が満たされると（ステップS7のYes）、事前学習を終了する。

[0046] 操作推論学習モデル70及び価値推論学習モデル80の、車両学習モデル60を操作aの実行対象として用いた事前学習が終了すると、学習部30は、車両学習モデル60に替えて、実車両2を操作aの実行対象として、操作推論学習モデル70及び価値推論学習モデル80を更に強化学習する（ステップS9）。

[0047] 次に、上記のドライvroボットを制御する操作推論学習モデルの学習システム及び学習方法の効果について説明する。

[0048] 本実施形態の学習制御システム10は、車速を含む車両2の走行状態sを基に、車両2を規定された指令車速に従って走行させるような、車両2の操作aを推論する操作推論学習モデル70と、車両2に搭載されて、操作aを基に車両2を走行させるドライvroボット（自動操縦ロボット）4を備え、操作推論学習モデル70を強化学習する、ドライvroボット4を制御する操作推論学習モデル70の学習システム10であって、車両2の実際の走行実績を基に車両2を模擬動作するように機械学習され、操作推論学習モデル70が推論した操作aを基に、車両2を模した走行状態sである模擬走行状態

○を出力する、車両学習モデル60を備え、車両学習モデル60が出力した模擬走行状態○を操作推論学習モデル70に適用することで、操作推論学習モデル70を事前に強化学習し、事前の強化学習が終了した後に、操作推論学習モデル70が推論した操作aを基に車両2を走行させて取得された走行状態sを操作推論学習モデル70に適用することで、操作推論学習モデル70を更に強化学習する。

また、本実施形態の学習制御方法は、車速を含む車両2の走行状態sを基に、車両2を規定された指令車速に従って走行させるような、車両2の操作aを推論する操作推論学習モデル70と、車両2に搭載されて、操作aを基に車両2を走行させるドライvroボット（自動操縦ロボット）4に関し、操作推論学習モデル70を強化学習する、ドライvroボット4を制御する操作推論学習モデル70の学習方法であって、車両2の実際の走行実績を基に車両2を模擬動作するように機械学習された車両学習モデル60により、操作推論学習モデル70が推論した操作aを基に、車両2を模した走行状態sである模擬走行状態○を出力し、模擬走行状態○を操作推論学習モデル70に適用することで、操作推論学習モデル70を事前に強化学習し、事前の強化学習が終了した後に、操作推論学習モデル70が推論した操作aを基に車両2を走行させて取得された走行状態sを操作推論学習モデル70に適用することで、操作推論学習モデル70を更に強化学習する。

強化学習により学習される操作推論学習モデル70は、強化学習の初期段階においては、例えばペダルを極端に高い頻度で操作するような、人間には不可能で、実車両に負担がかかる、好ましくない操作aを出力する可能性がある。

上記のような構成によれば、このような強化学習の初期段階においては、当該車両学習モデル60が、操作推論学習モデル70が推論した操作aを基に、車両2を模した走行状態sである模擬走行状態○を出力し、これを操作推論学習モデル70に適用することで、操作推論学習モデル70を事前に強化学習する。すなわち、強化学習の初期段階においては、実車両2を使用せ

ずに、操作推論学習モデル 70 を強化学習することができる。したがって、実車両 2 の負担を低減可能である。

また、事前学習が終了すると、実車両 2 を使用して操作推論学習モデル 70 を更に強化学習するため、車両学習モデル 60 のみを使用して操作推論学習モデル 70 を強化学習する場合に比べると、操作推論学習モデル 70 により出力する操作の学習精度を向上することができる。

[0049] 特に、上記のような構成においては、事前学習を、車両学習モデル 60 を操作 a の実行対象として行うため、事前学習の全過程において車両 2 を操作 a の実行対象とした場合に比べると、学習時間を低減可能である。

[0050] また、車両学習モデル 60 は、ニューラルネットワークで実現され、所定の時刻の走行実績が学習データとして、及び所定の時刻より後の時刻の走行状態が教師データとして入力され、後の時刻の模擬走行状態を出力してこれを教師データと比較することで機械学習される。

従来のように、車両モデルとして、車両の各構成要素に対して、動作を模した物理モデルを各々作成し、これらを組み合わせた物理モデルとして作成するのは、開発コストが嵩む。また、物理モデルの作成のためには実車両 2 の微細なパラメータや特性を熟知する必要があり、これらが入手できない場合においては、必要に応じて車両 2 を改造し、解析しなければならない。

上記のような構成によれば、車両学習モデル 60 がニューラルネットワークにより実現されているため、物理モデルとして車両学習モデル 60 を実現する場合に比べると実現が容易である。

また、車両学習モデル 60 は、操作推論学習モデル 70 の事前学習のみに使用され、事前学習後の強化学習には実車両 2 が用いられる。すなわち、操作推論学習モデル 70 の出力する操作 a の精度は、実車両 2 を操作 a の実行対象として用いた事前学習後の強化学習により高められるため、車両学習モデル 60 の車両 2 の模擬精度は過度に高くする必要がない。

以上が相乗し、学習制御システム 10 全体の開発が容易である。

[0051] また、走行状態 s は、車速に加え、アクセルペダル開度、ブレーキペダル

開度のいずれか、またはいずれかの組み合わせを含む。

上記のような構成によれば、上記のような学習制御システム 10 を適切に実現可能である。

[0052] なお、本発明のドライプロボットを制御する操作推論学習モデルの学習システム及び学習方法は、図面を参照して説明した上述の実施形態に限定されるものではなく、その技術的範囲において他の様々な変形例が考えられる。

[0053] 例えば、上記実施形態においては、事前に車両学習モデル 60 を操作 a の実行対象として操作推論学習モデル 70 を強化学習した後は、車両 2 を操作 a の実行対象として操作推論学習モデル 70 を強化学習していた。

事前学習後においては、操作推論学習モデル 70 が推論した操作により車両 2 を走行させることで、車両 2 の走行実績を更に取得することができる。この、新たに取得された走行実績を用いて、車両学習モデル 60 を更に学習し、模擬走行状態の推論精度を高めたうえで、事前学習後の強化学習において、推論した操作の実行及び走行状態の取得の対象として、車両 2 に加えて、更なる学習がなされた車両学習モデル 60 を併用することもできる。このような構成においては、車両 2 を使用して試験を行う時間が低減するため、操作推論学習モデル 70 の学習時間を低減可能である。

[0054] また、上記実施形態においては、車両学習モデル 60 の学習に使用する、車両 2 の実際の走行実績データを収集する際に、ドライプロボット 4 を使用した構成を説明したが、この場合に車両 2 を操縦するのはドライプロボット 4 に限られず、例えば人間であってもよい。この場合においては、上記実施形態で既に説明したように、アクセルペダル 2 c やブレーキペダル 2 d の操作量を計測するために、例えばカメラや赤外線センサが用いられ得る。

[0055] また、上記実施形態においては、走行状態として、車速、アクセルペダル開度、ブレーキペダル開度を用いていたが、これに限られない。例えば、走行状態は、車速に加え、アクセルペダル開度、ブレーキペダル開度、エンジン回転数、ギア状態、エンジン温度のいずれか、またはいずれかの組み合わせを含むように構成してもよい。

例えば、走行状態として、上記実施形態の構成に加えてエンジン回転数、ギア状態、エンジン温度を追加する場合においては、車両学習モデル60の入力には、車速系列i1、アクセルペダル系列i2、ブレーキペダル系列i3に加え、過去の時間におけるエンジン回転数の系列、ギア状態の系列、及びエンジン温度の系列が追加され得る。また、出力には、推定車速系列oに加え、将来の時間におけるエンジン回転数の系列、ギア状態の系列、及びエンジン温度の系列が追加され得る。

このように構成した場合においては、より精度の高い車両学習モデル60を生成することができる。

[0056] これ以外にも、本発明の主旨を逸脱しない限り、上記実施形態で挙げた構成を取捨選択したり、他の構成に適宜変更したりすることが可能である。

符号の説明

- [0057] 1 試験装置
- 2 車両
- 3 シャシーダイナモメータ
- 4 ドライブロボット（自動操縦ロボット）
- 10 学習制御システム（学習システム）
- 11 学習制御装置
- 20 ドライブロボット制御部
- 21 ペダル操作パターン生成部
- 22 車両操作制御部
- 23 駆動状態取得部
- 30 学習部
- 31 指令車速生成部
- 32 推論データ成形部
- 33 学習データ成形部
- 34 学習データ生成部
- 35 学習データ記憶部

- 4 0 強化学習部
- 4 1 操作内容推論部
- 4 2 状態行動価値推論部
- 4 3 報酬計算部
- 5 0 試験装置モデル
- 5 1 ドライブロボットモデル
- 5 2 車両モデル
- 5 3 シャシーダイナモメータモデル
- 6 0 車両学習モデル
- 7 0 操作推論学習モデル
- 8 0 価値推論学習モデル
- i 1 車速系列
- i 2 アクセルペダル系列
- i 3 ブレーキペダル系列
- a 操作
- s 走行状態
- o 模擬走行状態

請求の範囲

[請求項1] 車速を含む車両の走行状態を基に、前記車両を規定された指令車速に従って走行させるような、前記車両の操作を推論する操作推論学習モデルと、前記車両に搭載されて、前記操作を基に当該車両を走行させる自動操縦ロボットを備え、前記操作推論学習モデルを強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習システムであって、

前記車両の実際の走行実績を基に前記車両を模擬動作するように機械学習され、前記操作推論学習モデルが推論した前記操作を基に、前記車両を模した前記走行状態である模擬走行状態を出力する、車両学習モデルを備え、

前記車両学習モデルが出力した前記模擬走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを事前に強化学習し、当該事前の強化学習が終了した後に、前記操作推論学習モデルが推論した前記操作を基に前記車両を走行させて取得された前記走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを更に強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習システム。

[請求項2] 前記車両学習モデルは、ニューラルネットワークで実現され、所定の時刻を基点とした前記走行状態が学習データとして、及び前記所定の時刻より後の時刻の前記走行実績が教師データとして入力され、前記後の時刻の前記模擬走行状態を出力してこれを前記教師データと比較することで機械学習される、請求項1に記載の自動操縦ロボットを制御する操作推論学習モデルの学習システム。

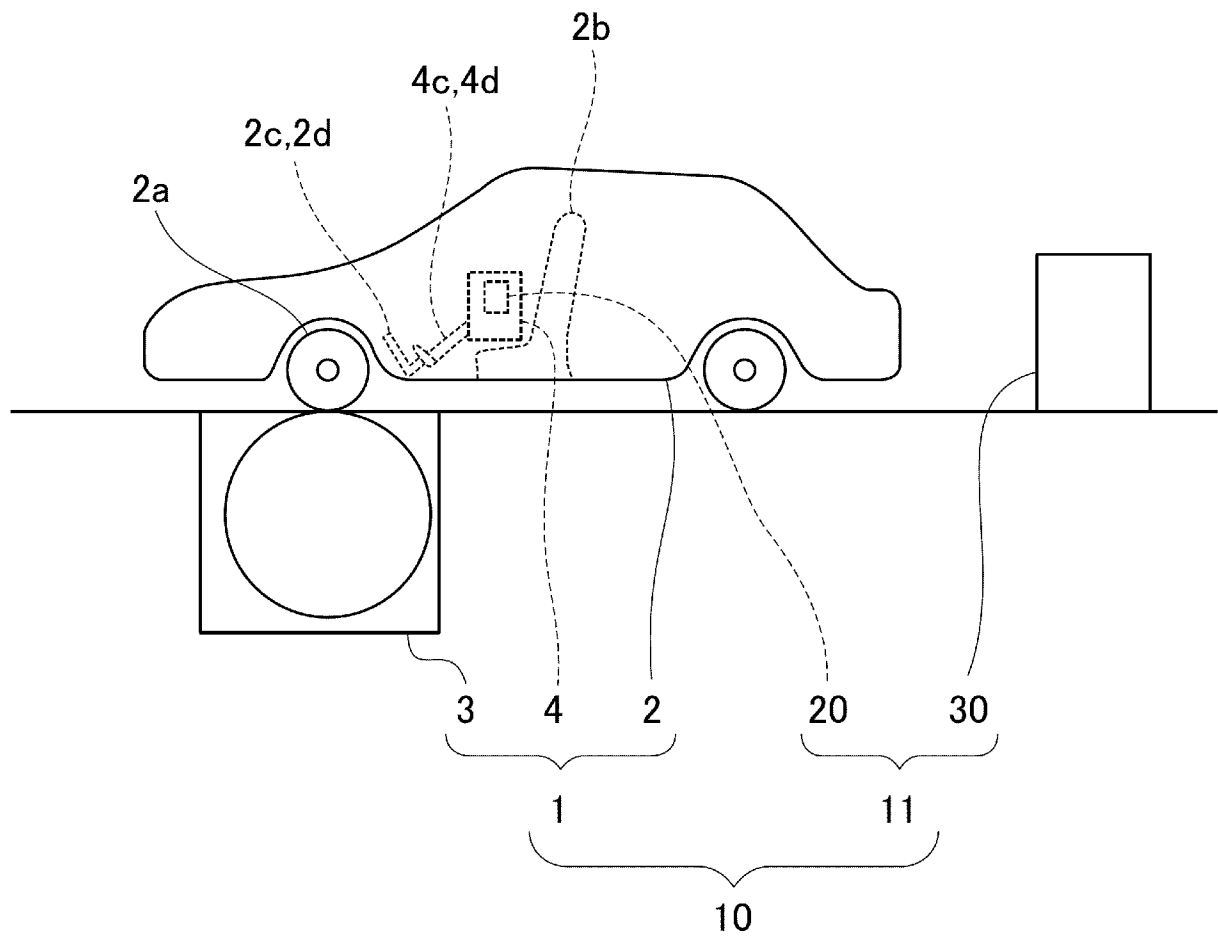
[請求項3] 前記走行状態は、前記車速に加え、アクセルペダル開度、ブレーキペダル開度、エンジン回転数、ギア状態、エンジン温度のいずれか、またはいずれかの組み合わせを含む、請求項1または2に記載の自動操縦ロボットを制御する操作推論学習モデルの学習システム。

[請求項4] 車速を含む車両の走行状態を基に、前記車両を規定された指令車速に従って走行させるような、前記車両の操作を推論する操作推論学習モデルと、前記車両に搭載されて、前記操作を基に当該車両を走行させる自動操縦ロボットに関し、前記操作推論学習モデルを強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習方法であって、

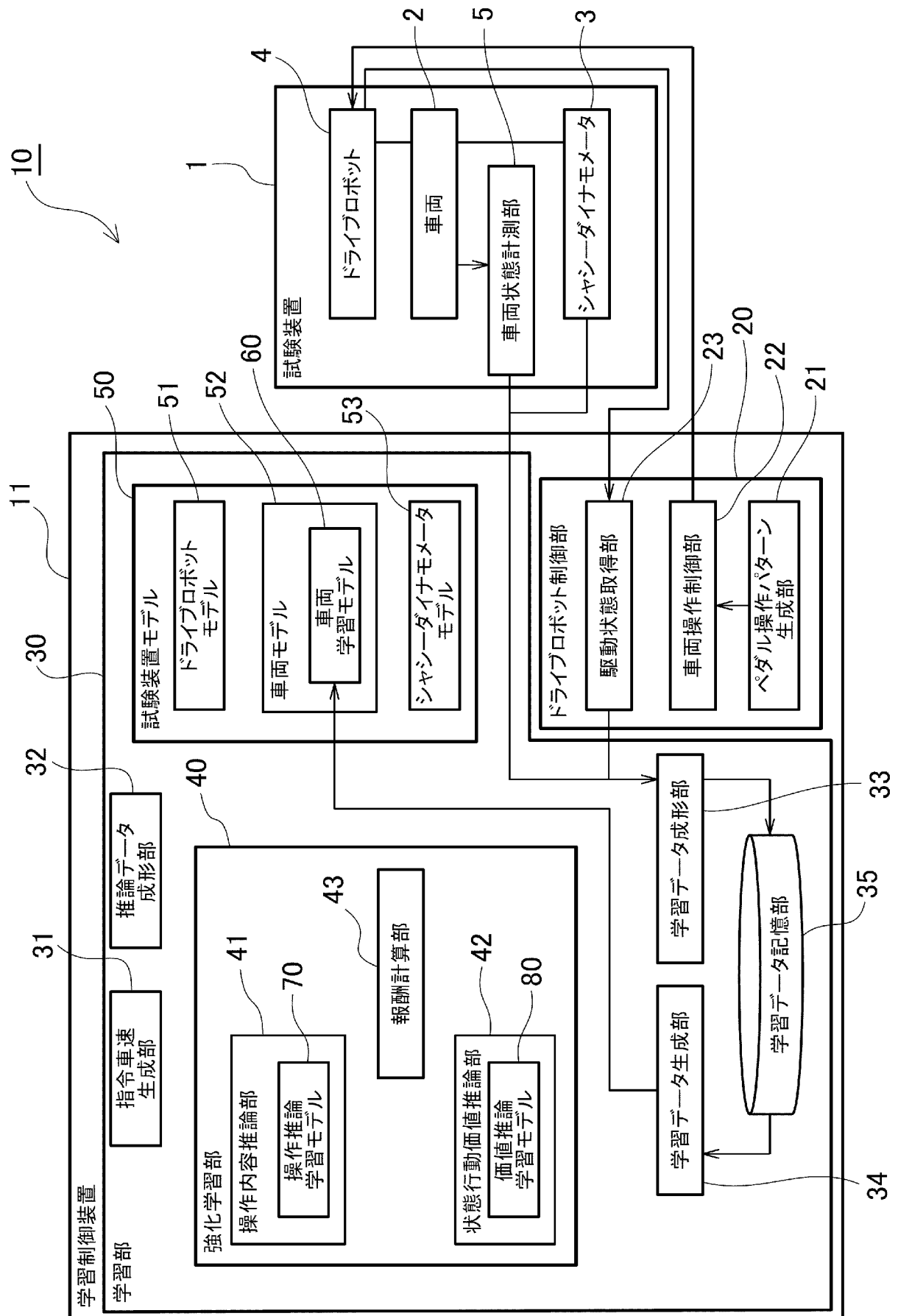
前記車両の実際の走行実績を基に前記車両を模擬動作するように機械学習された車両学習モデルにより、前記操作推論学習モデルが推論した前記操作を基に、前記車両を模した前記走行状態である模擬走行状態を出力し、当該模擬走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを事前に強化学習し、

前記事前の強化学習が終了した後に、前記操作推論学習モデルが推論した前記操作を基に前記車両を走行させて取得された前記走行状態を前記操作推論学習モデルに適用することで、前記操作推論学習モデルを更に強化学習する、自動操縦ロボットを制御する操作推論学習モデルの学習方法。

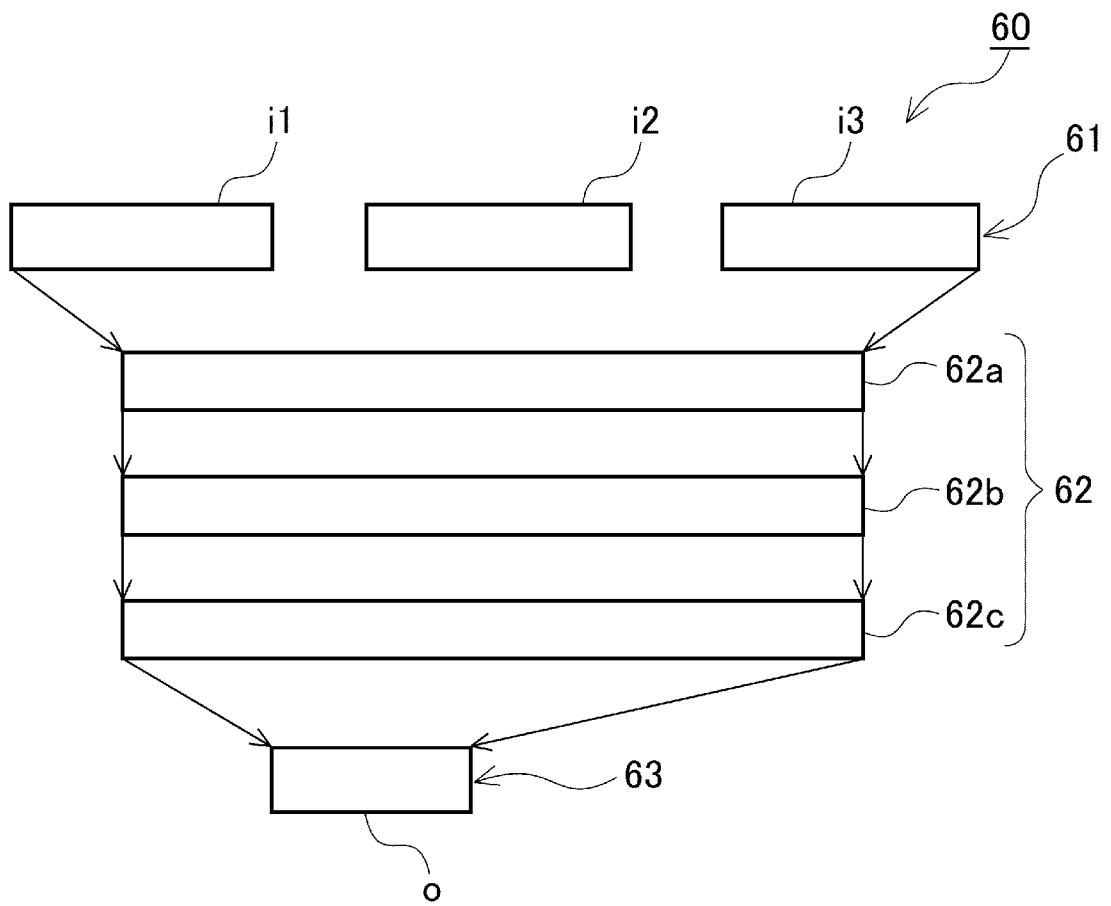
[図1]



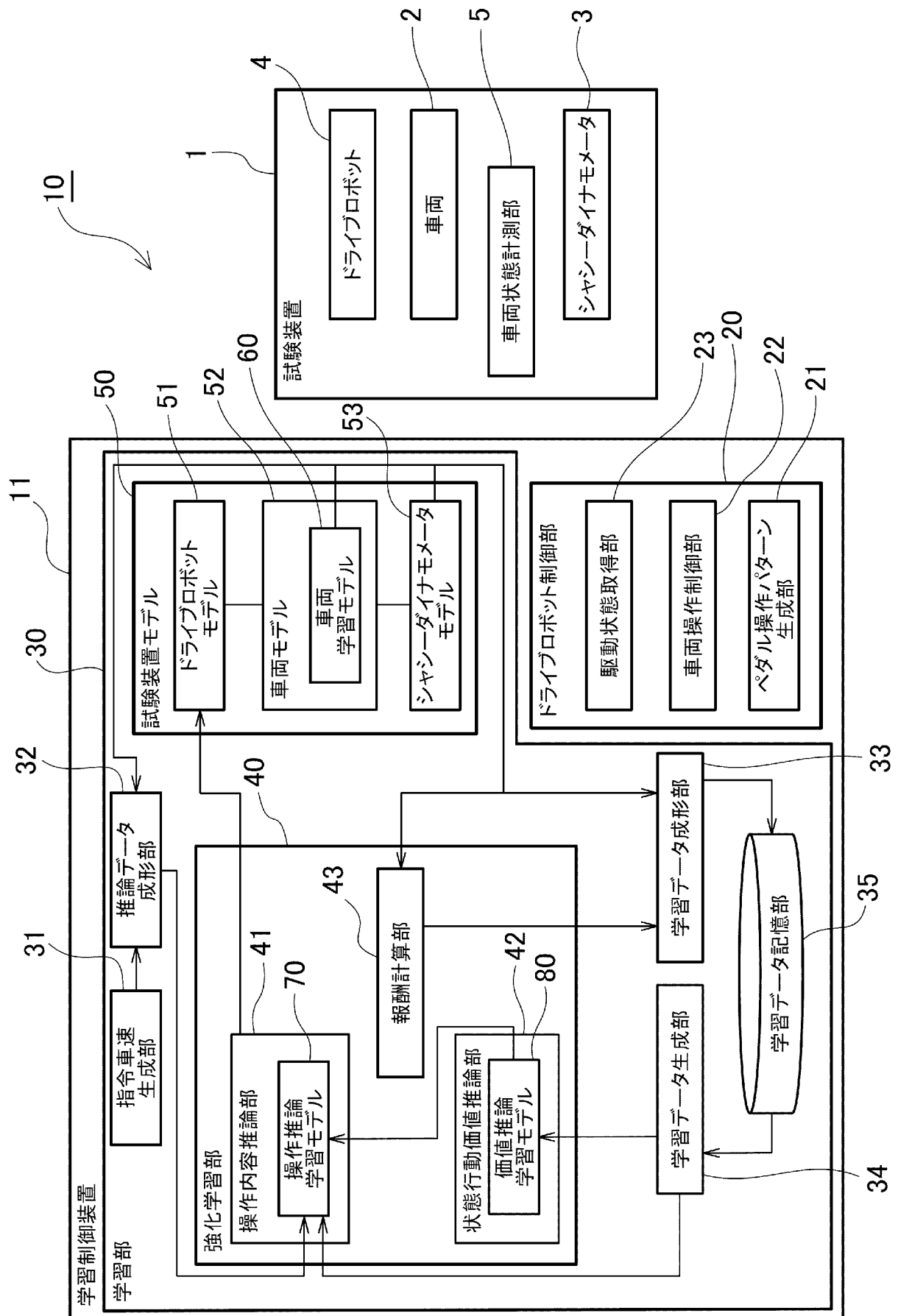
[図2]



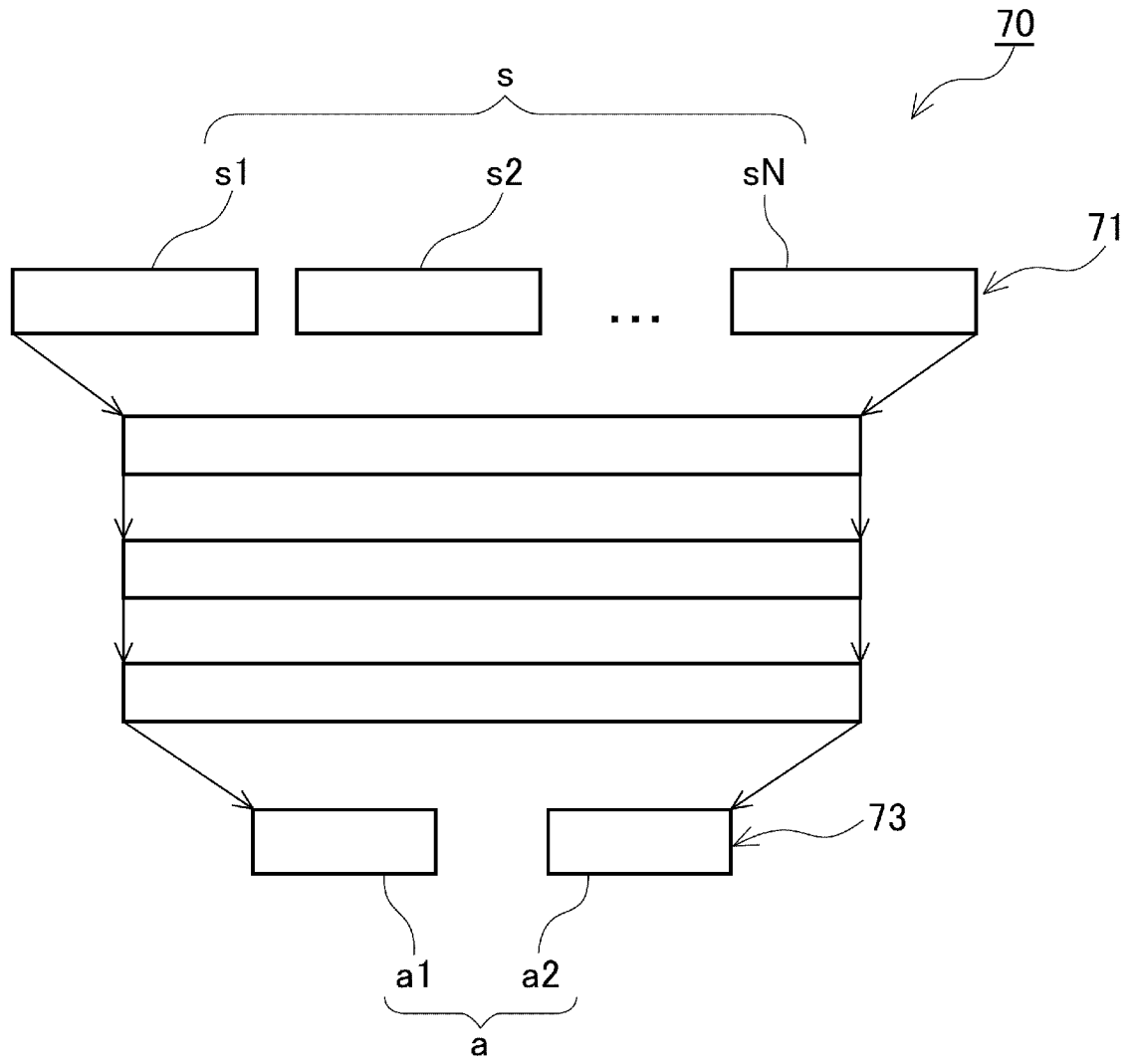
[図3]



[図4]



[図5]



[図6]

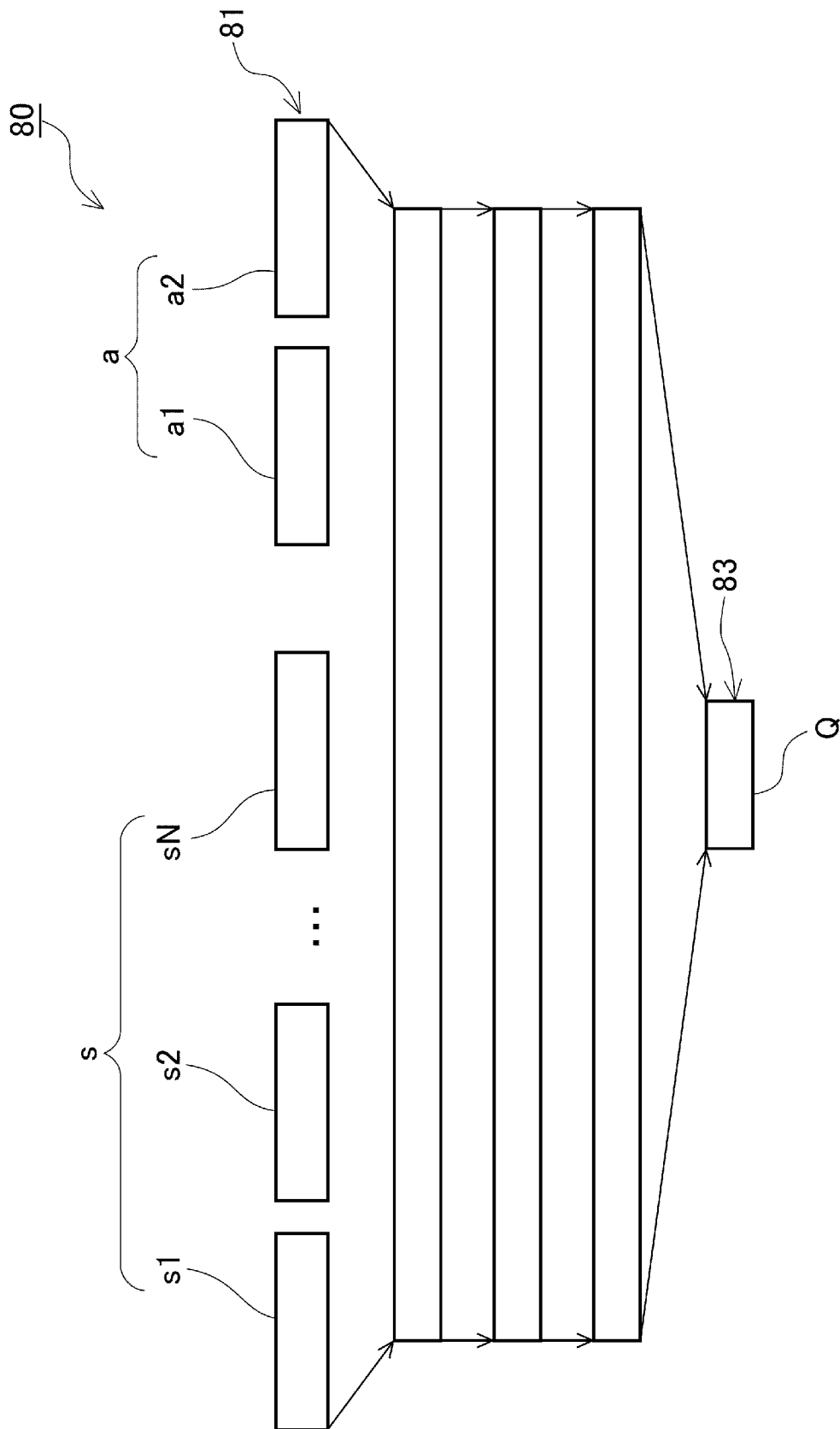
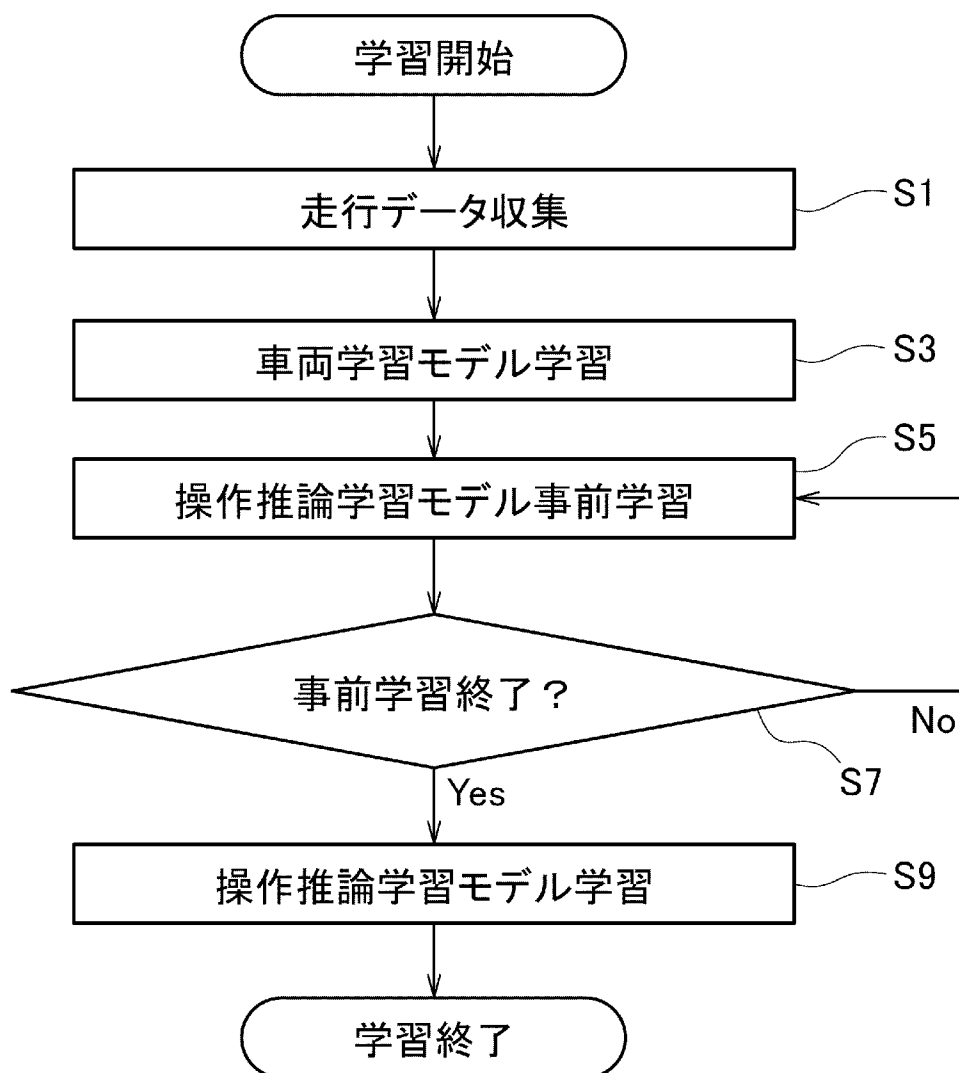


Figure 1 is a block diagram of a learning control device 10. The device is divided into a learning control section 11 and a test section 20. The learning control section 11 includes a learning unit 30, a reinforcement learning unit 40, and a driving robot control unit 50. The learning unit 30 contains a command speed generation unit 31, a data formation unit 32, and a data formation unit 33. The reinforcement learning unit 40 contains an operation content recommendation unit 41, a recommendation calculation unit 42, and a state action value recommendation unit 43. The driving robot control unit 50 contains a driving robot model 51, a vehicle model 52, and a vehicle status measurement model 53. The driving robot 21 contains a driving robot model 21, a vehicle model 22, and a vehicle status measurement unit 23. The vehicle 22 contains a vehicle status measurement unit 23. The vehicle status measurement unit 23 contains a driving robot model 23, a vehicle model 24, and a vehicle status measurement model 25.

[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/050747

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G01M17/007 (2006.01) i, G06N3/02 (2006.01) i, G06N20/00 (2019.01) i
FI: G01M17/007 D, G06N3/02, G06N20/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G01M17/007, G06N3/02, G06N20/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2020
Registered utility model specifications of Japan 1996-2020
Published registered utility model applications of Japan 1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-101025 A (FANUC LTD.) 28 June 2018, paragraphs [0091]-[0094], fig. 11	1-4
A	JP 2008-129527 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 05 June 2008, paragraphs [0008]-[0025], fig. 1, 2	1-4
A	JP 2019-29856 A (FANUC LTD.) 21 February 2019, paragraphs [0014]-[0059], fig. 1-6	1-4

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
26.02.2020

Date of mailing of the international search report
10.03.2020

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2019/050747

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
JP 2018-101025 A	28.06.2018	US 2018/0172983 A1 paragraphs [0104]- [0107], fig. 11 DE 102017011463 A1 CN 108205180 A	
JP 2008-129527 A	05.06.2008	(Family: none)	
JP 2019-29856 A	21.02.2019	US 2019/0036594 A1 paragraphs [0020]- [0068], fig. 1-6 DE 102018006044 A1 CN 109324536 A	

A. 発明の属する分野の分類（国際特許分類（IPC））

G01M 17/007(2006.01)i; G06N 3/02(2006.01)i; G06N 20/00(2019.01)i
FI: G01M17/007 D; G06N3/02; G06N20/00

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

G01M17/007; G06N3/02; G06N20/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922 - 1996年
日本国公開実用新案公報	1971 - 2020年
日本国実用新案登録公報	1996 - 2020年
日本国登録実用新案公報	1994 - 2020年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2018-101025 A（ファナック株式会社）28.06.2018（2018 - 06 - 28） [0091]-[0094], [図11]	1-4
A	JP 2008-129527 A（日本電信電話株式会社）05.06.2008（2008 - 06 - 05） [0008]-[0025], [図1]-[図2]	1-4
A	JP 2019-29856 A（ファナック株式会社）21.02.2019（2019 - 02 - 21） [0014]-[0059], [図1]-[図6]	1-4

☐ C欄の続きにも文献が列挙されている。

☒ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

“A” 特に関連のある文献ではなく、一般的技術水準を示すもの

“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

“O” 口頭による開示、使用、展示等に言及する文献

“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献

“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの

“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

“G” 同一パテントファミリー文献

国際調査を完了した日

26.02.2020

国際調査報告の発送日

10.03.2020

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目4番3号

権限のある職員（特許庁審査官）

川瀬 正巳 2J 5260

電話番号 03-3581-1101 内線 3252

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2019/050747

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-101025	A	28.06.2018	US 2018/0172983 A1 [0104]-[0107], Fig. 11 DE 102017011463 A1 CN 108205180 A	
JP	2008-129527	A	05.06.2008	(ファミリーなし)	
JP	2019-29856	A	21.02.2019	US 2019/0036594 A1 [0020]-[0068], FIGS1-6 DE 102018006044 A1 CN 109324536 A	