(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年3月24日(24.03.2022)

WIPO | PCT

(10) 国際公開番号
WO 2022/059484 A1

(51) 国際特許分類:
F02D 29/02 (2006.01)　*G01M 17/007* (2006.01)

(21) 国際出願番号：　PCT/JP2021/032055

(22) 国際出願日：　2021年9月1日(01.09.2021)

(25) 国際出願の言語：　日本語

(26) 国際公開の言語：　日本語

(30) 優先権データ：
特願 2020-154225　2020年9月15日(15.09.2020)　JP

(71) 出願人: 株式会社明電舎 (**MEIDENSHA CORPORATION**) [JP/JP]; 〒1416029 東京都品川区大崎２丁目１番１号 Tokyo (JP).

(72) 発明者: 金刺　泰宏 (**KANAZASHI, Yasuhiro**); 〒1416029 東京都品川区大崎２丁目１番１号 株式会社明電舎内 Tokyo (JP). 吉田　健人 (**YOSHIDA, Kento**); 〒1416029 東京都品川区大崎２丁目１番１号 株式会社明電舎内 Tokyo (JP).

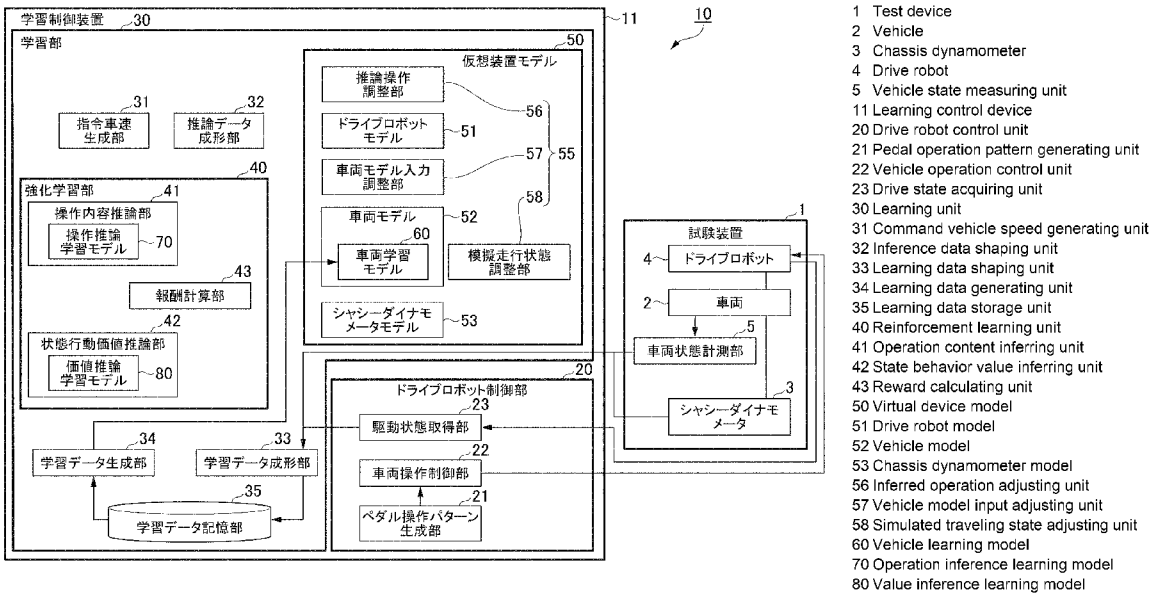
(74) 代理人: 園田・小林特許業務法人 (**SONODA & KOBAYASHI INTELLECTUAL PROPERTY LAW**); 〒1630434 東京都新宿区西新宿二丁目１番１号 新宿三井ビル３４階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) **Title:** LEARNING SYSTEM AND LEARNING METHOD FOR OPERATION INFERENCE LEARNING MODEL FOR CONTROLLING AUTOMATED DRIVING ROBOT

(54) 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

[図2]



1 Test device
2 Vehicle
3 Chassis dynamometer
4 Drive robot
5 Vehicle state measuring unit
11 Learning control device
20 Drive robot control unit
21 Pedal operation pattern generating unit
22 Vehicle operation control unit
23 Drive state acquiring unit
30 Learning unit
31 Command vehicle speed generating unit
32 Inference data shaping unit
33 Learning data shaping unit
34 Learning data generating unit
35 Learning data storage unit
40 Reinforcement learning unit
41 Operation content inferring unit
42 State behavior value inferring unit
43 Reward calculating unit
50 Virtual device model
51 Drive robot model
52 Vehicle model
53 Chassis dynamometer model
56 Inferred operation adjusting unit
57 Vehicle model input adjusting unit
58 Simulated traveling state adjusting unit
60 Vehicle learning model
70 Operation inference learning model
80 Value inference learning model

(57) **Abstract:** [Problem] To provide a learning system and a learning method for an operation inference learning model which controls an automated driving robot (drive robot), capable of easily suppressing a deterioration in the learning accuracy of the operation inference learning model resulting from differences in processing times between a vehicle model and an actual vehicle, when performing machine learning of the operation inference learning model with the vehicle model as the object of operation execution. [Solution] A learning system 10 for performing machine learning of an operation inference learning model 70 which controls an automated driving robot 4 is provided with a real environment 1 provided

WO 2022/059484 A1

[続葉有]

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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)）

with a vehicle 2 and the automated driving robot 4, mounted in the vehicle 2, and the operation inference learning model 70 for inferring operations of the vehicle 2 in such a way as to cause the vehicle 2 to travel in accordance with a prescribed command vehicle speed, on the basis of the traveling state of the vehicle 2 including the vehicle speed, wherein the automated driving robot 4 causes the vehicle 2 to travel on the basis of inferred operations inferred by the operation inference learning model 70, and wherein: the learning system 10 is provided with a virtual device model 50 provided with a vehicle model 52 which is set in such a way as to simulate the operations of the vehicle 2, and which outputs a simulated traveling state, being the traveling state imitating the vehicle 2, on the basis of the inferred operations; and the virtual device model 50 is provided with an adjusting unit 55 which adjusts the time from when the inferred operations are inferred until the same are used by the virtual device model 50, and from when the simulated traveling state is output until the same is applied to the operation inference learning model 70, on the basis of the actual time required from when the simulated traveling state is output by the vehicle model 52 on the basis of the inferred operations, upon input of the inferred operations, the operation inference learning model 70 is subjected to machine learning by the application of the simulated traveling state to the operation inference learning model 70, and the inferred operations are inferred in the real environment 1, until the inferred operations are input into the automated driving robot 4, the vehicle 2 is operated and travels, the traveling state of the vehicle 2 is acquired, and the traveling state is applied to the operation inference learning model 70.

(57) 要約：【課題】車両モデルを操作実行の対象として操作推論学習モデルを機械学習するに際し、車両モデルと実車両との処理時間の差異に起因する操作推論学習モデルの学習精度の低下を、容易に抑制可能な、自動操縦ロボット（ドライブロボット）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。【解決手段】車両２と、前記車両２に搭載された自動操縦ロボット４とを備える実環境１と、車速を含む前記車両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を備え、前記自動操縦ロボット４は、前記操作推論学習モデル７０が推論した推論操作を基に当該車両２を走行させ、前記操作推論学習モデル７０を機械学習する、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２を模擬動作するように設定され、前記推論操作を基に、前記車両２を模した前記走行状態である模擬走行状態を出力する、車両モデル５２を備えた、仮想装置モデル５０を備え、前記仮想装置モデル５０は、前記推論操作が入力されると、前記推論操作を基に、前記車両モデル５２により前記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７０に適用することで、前記操作推論学習モデル７０を機械学習し、前記実環境１において、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボット４に入力されて前記車両２が操作、走行され、前記車両２の前記走行状態が取得され、当該走行状態が前記操作推論学習モデル７０に適用されるまでに要する実時間を基に、前記推論操作が推論されてから前記仮想装置モデル５０において使用され、前記模擬走行状態が出力されてから前記操作推論学習モデル７０に適用されるまでの時間を調整する、調整部５５を備えている、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、所謂ドライブロボット（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、ドライブ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、特に近年においては、ドライブロボットを、車両の現在の状態を入力すると、車両を指令車速に従って走行させるような操作を推論するように機械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輌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、車輌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、車輌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、車両モデルを含めた仮想環境の、実車両を含めた実際の環境に対する再現精度が低いと、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。

仮想環境の再現精度向上に大きな影響を有する実際の環境の特性として、処理時間が挙げられる。

例えば、実車両と車両モデルとは、何らかの操作が入力されてから、これに対応して走行状態が変化するまでの時間に、差異がある。これら実車両と車両モデルに限らず、実際の環境と仮想環境に個別に、例えばドライブロボ

ットと、これを模擬動作するドライブロボットモデルのような、互いに対応する処理体系が設けられている場合には、これらの間には少なからず処理時間の際が生じ得る。

[0006] 実際の環境と仮想環境の間で異なる処理時間を有し得る他の例として、データ等の伝達が挙げられる。例えば、実際の環境においては、ドライブロボットはアクチュエータを制御することで実車両のアクセルペダルやブレーキペダルを、機械的に、直接操作する。これに対し、仮想環境においては、ドライブロボットモデルの出力を車両モデルに入力するのみであるため、これらの間に機械的な動作は存在しない。したがって、仮想環境におけるデータ等の伝達に要する時間は、実際の環境よりも短いものとなり得る。

ドライブロボットと実車両の間のみならず、例えば操作推論学習モデルとドライブロボットあるいはドライブロボットモデルの間や、実車両あるいは車両モデルと操作推論学習モデルの間でのデータ伝達に要する時間に関しても、これらは、実際の環境と仮想環境で異なる値となり得る。

このように、仮想環境で操作推論学習モデルが推論した操作をドライブロボットモデルに入力して車両モデルを操作し、操作後の走行状態である模擬走行状態を取得して操作推論学習モデルに適用するまでの処理時間は、実際の環境で操作推論学習モデルが推論した操作をドライブロボットに入力して実車両を操作し、操作後の実際の走行状態を取得して操作推論学習モデルに適用するまでの処理時間とは異なった、多くの場合においてはより短い時間となり得る。

[0007] ここで、例えば、処理時間が上記のように実際の環境よりも小さな値として設定された仮想環境を用いて、アクセルペダルの操作を推論するように学習された操作推論学習モデルが、実際の環境で、アクセルペダルを操作するために使用される場合を考える。

このような場合においては、操作推論学習モデルがアクセルペダルの操作を推論した後の、実際の試験環境における、実際の走行状態が取得されるまでの反応が、仮想環境の場合に比べると遅くなる。このため、実際の環境に

おいては、入力された操作に対応して正しく反応しようとしているにもかかわらず、操作推論学習モデルは実際の環境から想定された程度の十分な反応がないと認識する。結果として、操作推論学習モデルは、実際の環境に対し、より大きな反応を求めて、必要以上に大きくアクセルペダルを操作してしまう。

必要以上に大きな操作は、実車両に負担をかけるため、できるだけ出力しないように、低減するのが望ましい。

したがって、仮想環境においては、処理時間を、実際の環境にあわせて調整するように、構築することが必要となる。

[0008] 例えば車両モデルが機械学習モデルである場合には、上記のような、実際の環境と仮想環境の処理時間の差異が発覚した後に、車両モデルを、処理時間が実際の環境に適合するように再度学習することも考えられる。すなわち、この場合においては、車両モデルが実際の環境の処理時間をも含めて学習するため、処理時間の調整に要する手間は省くことができる。しかし、機械学習モデルの再度の学習には多くの計算時間を要し、容易に実行され得るものではなく、現実的ではない。

あるいは、仮想環境を、実環境全体の、遅延も含めた、総合的な動作を再現するように構築することも考えられる。しかし、この場合においては、例えば車種が変わる等の、実車両を含めた、試験の対象となる試験環境が部分的に変わるだけで、仮想環境全体の再度の学習が必要となる。このため、実現が容易ではない。

[0009] 本発明が解決しようとする課題は、車両モデルを操作実行の対象として操作推論学習モデルを機械学習するに際し、車両モデルと実車両との処理時間の差異に起因する操作推論学習モデルの学習精度の低下を、容易に抑制可能な、自動操縦ロボット（ドライブロボット）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

[0010] 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、車両と、前記車両に搭載された自動操縦ロボットとを備える実環境と、車速を含む前記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を備え、前記自動操縦ロボットは、前記操作推論学習モデルが推論した推論操作を基に当該車両を走行させ、前記操作推論学習モデルを機械学習する、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、前記車両を模擬動作するように設定され、前記推論操作を基に、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、車両モデルを備えた、仮想装置モデルを備え、前記仮想装置モデルは、前記推論操作が入力されると、当該推論操作を基に、前記車両モデルにより前記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機械学習し、前記実環境において、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボットに入力されて前記車両が操作、走行され、前記車両の前記走行状態が取得され、当該走行状態が前記操作推論学習モデルに適用されるまでに要する実時間を基に、前記推論操作が推論されてから前記仮想装置モデルにおいて使用され、前記模擬走行状態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、調整部を備えている、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。

[0011] また、本発明は、車両と、前記車両に搭載された自動操縦ロボットとを備える実環境と、車速を含む前記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に関し、前記自動操縦ロボットは、前記操作推論学習モデルが推論した推論操作を基に当該車両を走行させ、前記操作推論学習モデルを機械学習する、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、前記車両を模擬動作するように設定され、前記推論操作を基に、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、車両モデルを備えた、仮想装置モデルにおいて、前記推論操作が入力されると、当該推論

操作を基に、前記車両モデルにより前記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機械学習し、前記実環境において、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボットに入力されて前記車両が操作、走行され、前記車両の前記走行状態が取得され、当該走行状態が前記操作推論学習モデルに適用されるまでに要する実時間を基に、前記推論操作が推論されてから前記仮想装置モデルにおいて使用され、前記模擬走行状態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、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

[0012] 本発明によれば、車両モデルを操作実行の対象として操作推論学習モデルを機械学習するに際し、車両モデルと実車両との処理時間の差異に起因する操作推論学習モデルの学習精度の低下を、容易に抑制可能な、自動操縦ロボット（ドライブロボット）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

[0013] [図1]本発明の実施形態における、自動操縦ロボット（ドライブロボット）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6]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。

[図7]上記実施形態の第１変形例の学習システムの、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。

発明を実施するための形態

[0014] 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、ドライブロボット（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ブロボットと記載する。

[0015] 図１は、実施形態におけるドライブ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、及びドライブロボット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。

ドライブロボット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。ドライブロボット４は、第１アクチュエータ４ｃと第２アクチュエータ４ｄ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２ｃとブレーキペダル２ｄに当接するように設けられている。

[0016] ドライブロボット４は、後に詳説する学習制御装置１１によって制御されている。学習制御装置１１は、ドライブロボット４の第１アクチュエータ４ｃと第２アクチュエータ４ｄを制御することにより、車両２のアクセルペダル２ｃとブレーキペダル２ｄの開度を変更、調整する。

学習制御装置１１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１１は、車両２のアクセルペダル２ｃとブレーキペダル２ｄ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１１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。

[0017] 学習システム１０は、上記のような試験装置１と学習制御装置１１を備えている。

学習制御装置１１は、ドライブロボット制御部２０と学習部３０を備えている。

ドライブロボット制御部２０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３０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２０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３０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。

[0018] 図２は、学習システム１０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２ｃやブレーキペダル２ｄ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。

[0019] 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、及び仮想装置モデル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。仮想装置モデル５０は、ドライブロボットモデル（自動操縦ロボットモデル）５１、車両モデル５２、シャシーダイナモメータモデル５３、及び調整部５５を備えている。調整部５５は、推論操作調整部５６、車両モデル入力調整部５７、及び模擬走行状態調整部５８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。

仮想装置モデル５０の、車両モデル５２、ドライブロボットモデル５１及びシャシーダイナモメータモデル５３は、試験装置１の、車両２、ドライブロボット４、シャシーダイナモメータ３に対応して、これらを模擬動作するように構成された、例えばプログラムである。すなわち、仮想装置モデル５０は、物理的に現存する実際の環境である試験装置（実環境）１に対し、これを模擬動作するように構成された、仮想環境である。

[0020] 後に説明するように、操作内容推論部４１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４１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ブロボット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機械学習して学習モデル（操作推論学習モデル）７０を生成する。操作内容推論部４１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７０を使用して、車両２の操作を推論する。

すなわち、学習システム１０は大別して、機械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、学習システム１０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。

[0021] まず、操作の学習時における、学習制御装置１１の構成要素の挙動を説明する。

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、ドライブロボット制御部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。

ペダル操作パターン生成部２１は、ペダル２ｃ、２ｄ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、ＷＬＴＣ（Ｗｏｒｌｄｗｉｄｅ ｈａｒｍｏｎｉｚｅｄ Ｌｉｇｈｔ ｖｅｈｉｃｌｅｓ Ｔｅｓｔ Ｃｙｃｌｅ）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２１は、生成したペダル操作パターンを、車両操作制御部２２へ送信する。

[0022] 車両操作制御部２２は、ペダル操作パターン生成部２１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４ｃ、４ｄ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４ｃ、４ｄ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２３は、例えばアクチュエータ４ｃ、４ｄ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の走行状態が計測される。例えば、駆動状態取得部２３は上記のように、アクセルペダル２ｃの検出量と、ブレーキペダル２ｄ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３０の学習データ成形部３３へ送信される。

学習データ成形部３３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３５に保存する。

[0023] 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３４は学習データ記憶部３５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、仮想装置モデル５０に送信する。

学習部３０の、仮想装置モデル５０の車両モデル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。車両学習モデル６０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設定、本実施形態においては機械学習され、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ている。

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をともに、車両学習モデル６０と呼称する。

[0024] 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。

[0025] 本実施形態においては、車両学習モデル６０の入力は、任意の基準時刻を基点として、走行実績データ内の所定の第１時間だけ過去から基準時刻までの間の、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の入力は、基準時刻から所定の第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。これらアクセルペダル２ｃの操作量の系列、及びブレーキペダル２ｄの操作量の系列は、実際には、学習データ記憶部３５に保存された走行実績データ内の、基準時刻以降のアクセルペダル２ｃの検出量と、ブレーキペダル２ｄの検出量であり、これらが基準時刻において車両２に対して適用される操作として、車両学習モデル６０に入力される。

入力層６１は、上記のような車速の系列である車速系列ｉ１、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操作量系列ｉ２、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操作量系列ｉ３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ｉ１、ｉ２、ｉ３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ｉ１に対応する入力は、実際には、車速系列ｉ１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。

[0026] 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、基準時刻から所定の第３時間だけ将来の時刻（後の時刻）までの間の、推定された車速の系列である推定車速系列ｏ１と、アクセルペダル２ｃの検出量の系列であるアクセルペダル検出量系列ｏ２、及びブレーキペダル２ｄの検出量の系列であるブレーキペダル検出量系列ｏ３を含む、模擬走行状態ｏである。この、図３においては、一つの矩形として示されている模擬走行状態ｏの各々は、実際には、上記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。

[0027] 車両学習モデル６０においては、上記のように基準時刻の走行実績が入力されて、後の時刻の、車両２の走行を模した模擬走行状態ｏを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、基準時刻から第３時間だけ将来の時刻までの間の走行実績を、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が出力した模擬走行状態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、教師データと模擬走行状態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。

[0028] 車両学習モデル６０の学習が終了すると、学習システム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

係が示された学習システム１０のブロック図である。本実施形態においては、操作推論学習モデル７０は、強化学習により機械学習される。すなわち、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の出力した推論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。

[0029] 後に説明するように、操作内容推論部４１は、学習が中途段階の操作推論学習モデル７０によって、現時点から第１時間だけ将来の時刻までの間の車両２の操作を推論操作として出力し、これを推論操作調整部５６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、すなわちペダル操作量を出力する。

車両学習モデル６０の学習により、仮想装置モデル５０は、全体として、実環境としての試験装置１の各々を模擬動作するように構成されている。仮想装置モデル５０は、推論操作を受信する。

[0030] 推論操作調整部５６は、後に詳細に説明するように、推論操作が操作推論学習モデル７０によって推論されてからドライブロボットモデル５１に入力されるまでの時間を調整する。推論操作調整部５６は、時間が調整された推論操作を、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、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、推論操作調整部５６か

ら受信した、時間が調整された推論操作を基に、操作系の表現を車両２に対する実際のペダル操作量の値へと変換して、入力操作を生成する。より詳細には、ドライブロボットモデル５１は、入力操作としてペダル操作量の系列であるアクセルペダル操作量系列ｉ２とブレーキペダル操作量系列ｉ３を生成し、車両モデル入力調整部５７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、車両モデル入力調整部５７に送信する。

[0031] 車両モデル入力調整部５７は、後に詳細に説明するように、車速系列ｉ１と入力操作ｉ２、ｉ３が車両モデル５２に入力されるまでの時間を調整する。車両モデル入力調整部５７は、時間が調整された車速系列ｉ１と入力操作ｉ２、ｉ３を、車両モデル５２に送信する。

車両モデル５２は、車速系列ｉ１と入力操作ｉ２、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が模擬走行状態ｏを出力すると、車両モデル５２は模擬走行状態ｏをシャシーダイナモメータモデル５３と模擬走行状態調整部５８に送信する。

模擬走行状態調整部５８は、後に説明するように、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

模擬走行状態調整部５８によって時間が調整された模擬走行状態ｏは、推論データ成形部３２と強化学習部４０に送信される。

このように、仮想装置モデル５０は、推論操作が入力されると、推論操作を基に、車両モデル５２により模擬走行状態ｏを出力する。

[0032] 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、模擬走行状態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。

強化学習部４０は、模擬走行状態ｏと指令車速系列を、走行状態として操作内容推論部４１に送信する。

[0033] 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。

本実施形態においては、操作推論学習モデル７０は、走行状態の各々に対応する入力ノードを備えた入力層と、複数の中間層、及び複数の出力ノードを有する出力層を備えた、ニューラルネットワークである。

入力ノードの各々に、対応する走行状態の値が入力されると、重みを基にした演算がなされて、入力ノードの次の段として設けられた中間層の、中間ノードの各々に、演算結果が格納される。このような演算と、次の段の中間ノードへの演算結果の格納が、各中間層に対して順次実行される。最終的には、最終段の中間層内の中間ノードに格納された演算結果を基に、同様な演算がなされ、その結果が出力ノードに格納される。

操作推論学習モデル７０の出力ノードの各々は、操作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、操作推論学習モデル７０は、操作として、例えばアクセルペダル操作の系列とブレーキペダル操作の系列を推論する。

[0034] 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とブレーキペダル操作を、推論操作として仮想装置モデル５０に送信する。仮想装置モデル５０においては、推論操作調整部５６が推論操作の時間を調整し、ドライブロボットモデル５１がこれを基に入力操作となるアクセルペダル操作量系列ｉ２とブレーキペダル操作量系列ｉ３を生成する。そして、車両モデル入力調整部５７が入力操作の時間を調整したうえで車両学習モデル６０に送信する。車両学習モデル６０は、これを受信して、次の模擬走行状

態ｏを推論する。模擬走行状態調整部５８は模擬走行状態ｏの時間を調整する。このようにして、次の走行状態が生成される。

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。

[0035] 報酬計算部４３は、走行状態と、これに対応して操作推論学習モデル７０により推論された推論操作、及び当該推論操作を基に新たに生成された走行状態を基に、適切に設計された式により報酬を計算する。報酬は、推論操作、及びこれに伴う新たに生成された走行状態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推論操作を出力するように、強化学習が行われる。

報酬計算部４３は、走行状態、これに対応して推論された推論操作、当該推論操作を基に新たに生成された走行状態、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による推論操作の推論と、この推論操作に対応した、車両モデル５２による模擬走行状態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。

[0036] 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が推論した推論操作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推論操作を出力するように、強化学習を行う。行動価値は、走行状態と、これに対する推論操作を引数として、報酬が大きいほど行動価値を高くするように設計された関数として表わされる。本実施形態においては、この関数の計算を、走行状態と推論操作を入力として、行動価値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。

[0037] 操作学習データ生成部３４は、学習データ記憶部３５内の学習データを成形して、状態行動価値推論部４２へ送信する。

状態行動価値推論部４２は、成形された学習データを受信し、価値推論学習モデル８０を機械学習させる。

本実施形態においては、価値推論学習モデル８０は、走行状態と推論操作の各々に対応する入力ノードを備えた入力層と、複数の中間層、及び行動価値に対応する出力ノードを備えた、ニューラルネットワークである。価値推論学習モデル８０は、操作推論学習モデル７０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。

[0038] 状態行動価値推論部４２は、ＴＤ（Ｔｅｍｐｏｒａｌ Ｄｉｆｆｅｒｅｎｃｅ）誤差、すなわち、推論操作を実行する前の行動価値と、推論操作を実行した後の行動価値の誤差を小さくして、行動価値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推論操作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の学習が進むと、価値推論学習モデル８０は、より適切な行動価値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の値が学習前とは変わるため、これに伴い、行動

価値が高くなるような推論操作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、操作内容推論部４１は、例えば行動価値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、すなわち行動価値が大きくなるような推論操作を出力するように、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、出力される推論操作が変化するため、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。

[0039] 学習部３０は、この事前学習としての、車両学習モデル６０を推論操作の実行対象として用いた強化学習を、事前学習終了基準を満たすまで実行する。

[0040] 次に、上記のような車両学習モデル６０を用いた操作推論学習モデル７０の事前学習における、調整部５５の挙動を説明する。

操作推論学習モデル７０が精度よく学習されるためには、仮想装置モデル５０において、試験装置１の再現精度を高める必要がある。この際に特に重要となるのは、試験装置１における処理時間を、仮想装置モデル５０においても正確に再現することである。

仮想環境で、すなわち仮想装置モデル５０において、操作推論学習モデル７０が推論した推論操作をドライブロボットモデル５１に入力して車両モデル５２を操作し、操作後の走行状態である模擬走行状態ｏを取得して操作推論学習モデル７０に適用するまでの処理時間は、実際の環境で、すなわち試験装置１において、操作推論学習モデル７０が推論した推論操作をドライブ

ロボット４に入力して車両２を操作し、操作後の実際の走行状態を取得して操作推論学習モデル７０に適用するまでの処理時間とは異なった、多くの場合においてはより短い時間となり得る。

このため、操作推論学習モデル７０が例えばアクセルペダル２ｃの操作を推論した後の、実際の試験装置１における、実際の走行状態が取得されるまでの反応が、仮想装置モデル５０の場合に比べると遅くなることが想定される。これにより、特に本実施形態のように、操作推論学習モデル７０の学習段階に応じて、推論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へと変更するような場合には、次のような不都合が生じる。すなわち、試験装置１においては、入力された推論操作に対応して正しく反応しようとしているにもかかわらず、操作推論学習モデル７０は試験装置１から想定された程度の十分な反応がないと認識する。結果として、操作推論学習モデル７０は、試験装置１に対し、より大きな反応を求めて、必要以上に大きくアクセルペダルを操作するような、推論操作を推論してしまう。

このような、実際の試験装置１を用いた際に必要以上に大きな操作が推論されるのを抑制するために、調整部５５は、仮想装置モデル５０を用いて操作推論学習モデル７０を学習させる際に、推論操作や模擬走行状態ｏが使用される時間を調整する。

[0041] 後に説明するように、事前学習終了後に、試験装置１を用いて操作推論学習モデル７０を学習する際には、操作推論学習モデル７０の出力した推論操作は、試験装置１に送信されて、ドライブロボット４に入力される。このため、試験装置１を用いた場合には、推論操作が推論されてからドライブロボット４へ入力されるまでの、一定の、実環境での伝達時間が必要である。

同様に、事前学習時に、仮想装置モデル５０を用いて操作推論学習モデル７０を学習する際にも、操作推論学習モデル７０の出力した推論操作は、仮想装置モデル５０に送信されて、ドライブロボットモデル５１に入力される。このため、仮想装置モデル５０を用いた場合にも、推論操作が推論されて

からドライブロボットモデル５１へ入力されるまでの、一定の、仮想環境での伝達時間が必要である。

ここで、本実施形態においては、ドライブロボット４は、学習制御装置１１とは独立して別個に設けられた装置であるが、ドライブロボットモデル５１は、学習制御装置１１内に設けられた、学習制御装置１１の構成要素である。したがって、操作推論学習モデル７０によって推論された推論操作の、実環境での伝達時間は、仮想環境での伝達時間よりも長いものとなり得る。

[0042] 推論操作調整部５６は、推論操作の入力時における、仮想環境での伝達時間を、実環境での伝達時間と同等となるように調整する。すなわち、推論操作調整部５６は、試験装置１を用いた場合において、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されるまでの時間を基に、推論操作が推論されてからドライブロボットモデル５１に入力されるまでの時間を調整する。

より詳細には、推論操作調整部５６は、操作推論学習モデル７０が出力した推論操作が、ドライブロボットモデル５１に、推論操作が出力された時刻から上記の実環境での伝達時間後に、入力されるように、推論操作の伝達時間を遅延させる。

[0043] また、事前学習終了後に、試験装置１を用いて操作推論学習モデル７０を学習する際には、ドライブロボット４は、第１及び第２アクチュエータ４ｃ、４ｄを制御することで、車両２のアクセルペダル２ｃやブレーキペダル２ｄを、機械的に、直接操作する。このため、試験装置１を用いた場合には、推論操作がドライブロボット４へ入力されてから車両２が操作されるまでの、一定の、実環境での作動時間が必要である。

同様に、事前学習時に、仮想装置モデル５０を用いて操作推論学習モデル７０を学習する際にも、ドライブロボットモデル５１は、入力操作を車両モデル５２に入力することで、車両モデル５２を操作する。このため、仮想装置モデル５０を用いた場合にも、入力操作が車両モデル５２に入力されるまでの、一定の、仮想環境での作動時間が必要である。

ここで、本実施形態においては、ドライブロボット４は上記のように、車両２を機械的に操作するのに対し、ドライブロボットモデル５１と車両モデル５２は、同一の学習制御装置１１内に設けられた、プログラムなどの構成要素である。したがって、車両２の操作に要する、実環境での作動時間は、車両モデル５２の操作に要する、仮想環境での作動時間よりも長いものとなり得る。

[0044] 車両モデル入力調整部５７は、仮想環境での作動時間を、実環境での作動時間と同等となるように調整する。すなわち、車両モデル入力調整部５７は、試験装置１を用いた場合において、推論操作がドライブロボット４に入力されてから車両２が操作されるまでの時間を基に、入力操作が車両モデル５２に入力されるまでの時間を調整する。

より詳細には、車両モデル入力調整部５７は、ドライブロボットモデル５１が出力した入力操作が、車両モデル５２に、入力操作が出力された時刻から上記の実環境での作動時間後に、入力されるように、入力操作の伝達時間を遅延させる。

[0045] 更に、事前学習終了後に、試験装置１を用いて操作推論学習モデル７０を学習する際には、車両２がドライブロボット４によって操作、走行された結果として、車両状態計測部５によって走行状態が取得され、学習部３０に送信されて加工され、操作推論学習モデル７０に入力される。このため、試験装置１を用いた場合には、走行状態が取得されてから操作推論学習モデル７０へ適用されるまでの、一定の、実環境での伝達時間が必要である。

同様に、事前学習時に、仮想装置モデル５０を用いて操作推論学習モデル７０を学習する際にも、車両モデル５２がドライブロボットモデル５１によって操作、走行された結果として、模擬走行状態ｏが取得され、最終的には操作推論学習モデル７０に入力される。このため、仮想装置モデル５０を用いた場合にも、模擬走行状態ｏが取得されてから操作推論学習モデル７０へ適用されるまでの、一定の、仮想環境での伝達時間が必要である。

ここで、本実施形態においては、車両２や車両２の走行状態を取得する車

両状態計測部５は、学習制御装置１１とは独立して別個に設けられた装置であるが、車両モデル５２は、学習制御装置１１内に設けられた、学習制御装置１１の構成要素である。したがって、走行状態が操作推論学習モデル７０に適用されるまでの、実環境での伝達時間は、模擬走行状態ｏが操作推論学習モデル７０に適用されるまでの、仮想環境での伝達時間よりも、長いものとなり得る。

[0046] 模擬走行状態調整部５８は、模擬走行状態ｏの出力時における、仮想環境での伝達時間を、走行状態の実環境での伝達時間と同等となるように調整する。すなわち、模擬走行状態調整部５８は、試験装置１を用いた場合において、車両２の走行状態が取得されてから操作推論学習モデル７０に適用されるまでの時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

より詳細には、模擬走行状態調整部５８は、車両モデル５２が出力した模擬走行状態ｏが、操作推論学習モデル７０に、模擬走行状態ｏが出力された時刻から上記の実環境での伝達時間後に、入力されるように、模擬走行状態ｏの伝達時間を遅延させる。

[0047] このように、調整部５５は、実環境すなわち試験装置１において、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、当該走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、推論操作が推論されてから仮想装置モデル５０において使用され、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

[0048] 上記のような時間の調整は、より詳細には、例えば次のように、推論操作や模擬走行状態ｏ等のデータの伝達を遅延させることで行われる。ここでは、模擬走行状態調整部５８を例として説明するが、推論操作調整部５６、車両モデル入力調整部５７も同様に構成することが可能である。

模擬走行状態調整部５８は、リングメモリを備えている。リングメモリは、例えば、メモリ上で一定の長さの配列を確保し、当該配列の末尾を超えて

アクセスがなされた際には、当該配列の先頭をアクセスするように構成されている。

模擬走行状態調整部５８は、車両モデル５２が模擬走行状態ｏを推論するたびに、リングメモリの先頭位置に模擬走行状態ｏを格納し、先頭位置を新たに格納された模擬走行状態ｏの先に移動させる。模擬走行状態ｏは車両モデル５２によって続々と推論されるが、格納対象がリングメモリであるため、これを適切に設計した際には、メモリの末尾を意識することなく、データの格納が可能である。

[0049] ここで、データの伝達を遅延させる時間をT_{delay}、車両モデル５２において模擬走行状態ｏの推論がなされる時間間隔をT_{sim}、リングメモリに格納される模擬走行状態ｏのデータサイズをL_{state}とすると、これらの値と、T_{delay}時間の間にリングメモリに格納される模擬走行状態ｏの総データサイズL_{num}との間には、次の関係が成立する。

$$T_{delay} = (L_{num} / L_{state}) \times T_{sim}$$

すなわち、模擬走行状態調整部５８は、リングメモリから、データサイズL_{num}だけ先頭位置から後方に位置して格納された模擬走行状態ｏを取得し、これを模擬走行状態ｏとして出力することで、模擬走行状態ｏの出力を時間T_{delay}だけ遅延させている。

ここで、リングメモリの大きさを、上記の総データサイズL_{num}と一致させると、データサイズL_{num}だけ先頭位置から後方に位置して格納された模擬走行状態ｏを取得する際に、実際には、リングメモリの先頭位置に現在格納されている模擬走行状態ｏを取得すればよいので、好適である。すなわち、この場合には、模擬走行状態調整部５８の実装が容易となる。

[0050] 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を推論操作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を推論操作の実行対象として、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５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

示された学習システム１０のブロック図である。

[0051] 操作内容推論部４１は、現時点から第１時間だけ将来の時刻までの間の車両２の推論操作を出力し、これを車両操作制御部２２に送信する。

車両操作制御部２２は、受信した推論操作を、ドライブロボット４の第１及び第２アクチュエータ４ｃ、４ｄ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４ｃ、４ｄ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と車両状態計測部５は、車両２の車速、アクセルペダル２ｃとブレーキペダル２ｄの操作量を検出して各々の系列を生成し、推論データ成形部３２に送信する。

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、各系列を受信し、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。

[0052] 強化学習部４０は、試験装置モデル５０により生成される推定車速系列ｏの替わりに上記の各系列を用いて、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を推論操作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３５に蓄積する。強化学習部４０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８０を学習し、その後操作推論学習モデル７０を学習する。

学習部３０は、学習データの蓄積と、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。

[0053] 学習部３０は、車両２を推論操作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。

[0054] 次に、車両２の性能測定に際して推論操作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、学習システム

１０の各構成要素の挙動について説明する。

[0055] 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の推論操作を推論する。

操作内容推論部４１は、推論した推論操作を、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から推論操作を受信し、この推論操作に基づき、ドライブロボット４を操作する。

[0056] 次に、図１～図５、及び図６を用いて、上記の学習システム１０を用いた、ドライブロボット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６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、ドライブロボット制御部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）。

[0057] 車両学習モデル６０の学習が終了すると、学習システム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。この際には、調整部５５は、実環境すなわち試験装置１において、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、当該走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、推論操作が推論されてから仮想装置モデル５０において使用され、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

学習部３０は、この事前学習としての、車両学習モデル６０を推論操作の実行対象として用いた強化学習を、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。

[0058] 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を推論操作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を推論操作の実行対象として、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）。

[0059] 次に、上記のドライブ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の効果について説明する。

[0060] 本実施形態の学習システム１０は、車両２と、車両２に搭載されたドライブロボット（自動操縦ロボット）４とを備える試験装置（実環境）１と、車速を含む車両２の走行状態を基に、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の操作を推論する操作推論学習モデル７０を備え、ドライブロボット４は、操作推論学習モデル７０が推論した推論操作を基に車両２を走行させ、操作推論学習モデル７０を機械学習する、ドライブロボ

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、車両２を模擬動作するように設定され、推論操作を基に、車両２を模した走行状態である模擬走行状態ｏを出力する、車両モデル５２を備えた、仮想装置モデル５０を備え、仮想装置モデル５０は、推論操作が入力されると、推論操作を基に、車両モデル５２により模擬走行状態ｏを出力し、当該模擬走行状態ｏ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を機械学習し、試験装置１において、推論操作が推論されてから、推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、推論操作が推論されてから仮想装置モデル５０において使用され、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する、調整部５５を備えている。

また、本実施形態の学習制御方法は、車両２と、車両２に搭載されたドライブロボット（自動操縦ロボット）４とを備える試験装置（実環境）１と、車速を含む車両２の走行状態を基に、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の操作を推論する操作推論学習モデル７０とに関し、ドライブロボット４は、操作推論学習モデル７０が推論した推論操作を基に車両２を走行させ、操作推論学習モデル７０を機械学習する、ドライブロボット４を制御する操作推論学習モデル７０の学習方法であって、車両２を模擬動作するように設定され、推論操作を基に、車両２を模した走行状態である模擬走行状態ｏを出力する、車両モデル５２を備えた、仮想装置モデル５０において、推論操作が入力されると、推論操作を基に、車両モデル５２により模擬走行状態ｏを出力し、当該模擬走行状態ｏ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を機械学習し、試験装置１において、推論操作が推論されてから、推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、推論操作が推論されてから仮想装置モデル５０において使用され、模擬走行状態

ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

上記のような構成によれば、既に説明したように、試験装置１における、データの伝達に要する遅延時間や機械的な動作時間が、仮想装置モデル５０において考慮され、結果として、試験装置１における処理時間が、仮想装置モデル５０においても正確に再現される。このため、仮想装置モデル５０における、入力された推論操作に対する反応を、試験装置１における反応に一致させるに際し、少なくとも、試験装置１と仮想装置モデル５０の間の処理時間による影響は低減される。

更に、試験装置１と仮想装置モデル５０の間の処理時間による影響が低減されるため、車両モデル５２を学習させた後に、試験装置１と仮想装置モデル５０の処理時間の差異が発覚し、車両モデル５２を処理時間が試験装置１に適合するように再度学習するという事態の発生も抑制される。このため、実現が容易である。

したがって、車両モデル５２を操作実行の対象として操作推論学習モデル７０を機械学習するに際し、車両モデル５２と実車両２との処理時間の差異に起因する操作推論学習モデル７０の学習精度の低下を、容易に抑制可能である。

[0061] 特に、処理時間が例えば実際の環境よりも小さな値として設定された仮想環境を用いて、アクセルペダルの操作を推論するように学習された操作推論学習モデルが、実際の環境で、アクセルペダルを操作するために使用される場合においては、操作推論学習モデルがアクセルペダルの操作を推論した後の、実際の試験環境における、実際の走行状態が取得されるまでの反応が、仮想環境の場合に比べると遅くなる。このため、実際の環境においては、入力された操作に対応して正しく反応しようとしているにもかかわらず、操作推論学習モデルは実際の環境から想定された程度の十分な反応がないと認識する。結果として、操作推論学習モデルは、実際の環境に対し、より大きな反応を求めて、必要以上に大きくアクセルペダルを操作してしまう。

これに対し、本実施形態においては、上記のように、試験装置１と仮想装置モデル５０の間の処理時間による影響が低減されるため、上記のような必要以上に大きな操作を抑制し、実車両への負担を低減可能である。

[0062] また、仮想装置モデル５０は、ドライブロボット４を模擬動作するように設定され、推論操作を入力として、車両モデル５２へ入力される入力操作を出力する、ドライブロボットモデル（自動操縦ロボットモデル）５１を更に備え、調整部５５は、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されるまでの時間を基に、推論操作が推論されてからドライブロボットモデル５１に入力されるまでの時間を調整する、推論操作調整部５６と、推論操作がドライブロボット４に入力されてから車両２が操作されるまでの時間を基に、入力操作が車両モデル５２に入力されるまでの時間を調整する、車両モデル入力調整部５７と、車両２の走行状態が取得されてから操作推論学習モデル７０に適用されるまでの時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する、模擬走行状態調整部５８と、を備えている。

また、操作は、アクセルペダル２ｄとブレーキペダル２ｄのいずれか一方または双方のペダルの、ペダル操作量を含む。

また、車両モデル５２は、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、推論操作を基に模擬走行状態ｏを出力する、車両学習モデル６０を備えている。

特に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ている。

上記のような構成によれば、学習システム１０を適切に実現可能である。

[0063] また、操作推論学習モデル７０は、強化学習されている。

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２ｃ、２ｄ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、好ましくない推論操作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が推論した推論操作を基に、車両２を模した走行状態ｓである模擬走行状態ｏ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。

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を推論操作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を推論操作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。

[0064] ［実施形態の第１変形例］

次に、上記実施形態として示したドライブ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の第１変形例を説明する。図７は、本第１変形例における学習システムの、操作推論学習モデル７０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。本第１変形例における学習システムは、上記実施形態の学習システム１０とは、調整部５５Ａが推論操作調整部５６と車両モデル入力調整部５７を備えておらず、模擬走行状態調整部５８Ａのみを備えている点が異なっている。

[0065] 本変形例における模擬走行状態調整部５８Ａは、上記実施形態における推論操作調整部５６、車両モデル入力調整部５７、及び模擬走行状態調整部５８の各々において調整された、全ての時間を調整する。

すなわち、本変形例における模擬走行状態調整部５８Ａは、試験装置１において、推論操作が推論されてから、推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、走行状態が

操作推論学習モデル７０に適用されるまでに要する実時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

このように、本変形例における模擬走行状態調整部５８Ａは、仮想装置モデル５０を、試験装置１に対応する１つの制御処理系と見做して、仮想装置モデル５０全体における時間をまとめて調整している。

[0066] 上記実施形態においては、試験装置１において、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されるまでの時間、推論操作がドライブロボット４に入力されてから車両２が操作されるまでの時間、及び車両２の走行状態が取得されてから操作推論学習モデル７０に適用されるまでの時間の各々を測定し、測定結果を推論操作調整部５６、車両モデル入力調整部５７、及び模擬走行状態調整部５８の各々に、個別に反映させる必要があった。

これに対し、本変形例においては、上記のように仮想装置モデル５０全体の処理時間を測定し、測定結果を模擬走行状態調整部５８Ａに反映させればよいため、仮想装置モデル５０の構築が容易である。

本変形例が、既に説明した実施形態と同様な他の効果を奏することは言うまでもない。

[0067] ［実施形態の第２変形例］

次に、上記実施形態として示したドライブ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の第２変形例を説明する。本第２変形例における学習システムは、上記実施形態の学習システム１０とは、調整部５５が、推論操作の分解能を、仮想装置モデル５０における分解能へと変換し、模擬走行状態ｏの分解能を、車両２に対して取得される走行状態の分解能へと変換する点が異なっている。

[0068] 上記実施形態においては、推論操作調整部５６は、推論操作が推論されてから、当該推論操作がドライブロボット４に入力されるまでの時間を基に、推論操作が推論されてからドライブロボットモデル５１に入力されるまでの

時間を調整していた。本変形例においては、推論操作調整部５６は、これに加えて、操作推論学習モデル７０が推論した推論操作の分解能を、仮想装置モデル５０における、例えば通信パケットのデータサイズに応じた、分解能へと変換する。

また、上記実施形態においては、模擬走行状態調整部５８は、車両２の走行状態が取得されてから操作推論学習モデル７０に適用されるまでの時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整していた。本変形例においては、模擬走行状態調整部５８は、これに加えて、模擬走行状態ｏの分解能を、試験装置１における、例えばセンサや通信パケットのデータサイズに応じた、分解能へと変換する。

[0069] 上記のような時間の調整は、より詳細には、例えば次のように行われる。

例えば仮想装置モデル５０において、プログラムによって、試験装置１の各々の物理動作が記述され、再現される場合においては、各変数のデータ型としては、浮動小数点が使用される。推論操作調整部５６においては、仮想装置モデル５０に入力されるデータの型を浮動小数点へと変換する。逆に、模擬走行状態調整部５８においては、仮想装置モデル５０から出力するデータの型を浮動小数点から他の型へと変換する。

例えば、アクセルペダル２ｃに対する操作を想定した場合に、分解能を変換する前のペダル開度をP_{base}、変換後の１ビット分解能をB_{real}、分解能を変換した後の値をP_{chg}、浮動小数点の変数を整数へと変換する関数をＩｎｔｅｇｅｒ（）とすると、模擬走行状態調整部５８においてデータの型を浮動小数点から他の型へと変換する場合に、次の式が適用可能である。

$$P_{chg} = Integer(P_{base} / B_{real}) \times B_{real}$$

[0070] 上記実施形態においては、分解能が高い操作推論学習モデル７０が微小な推論操作を出力した場合、同等の分解能を有する仮想装置モデル５０はこれに対応し反応することが可能であるため、この微小な推論操作が有効なものであると操作推論学習モデル７０が学習する。このように学習された操作推論学習モデル７０が、実際の試験装置１に対して推論操作を出力する場合に

は、試験装置１の分解能が低いと、操作推論学習モデル７０が出力する微小な推論操作が試験装置１に効果的に反映されない。このため、試験装置１における推論操作の反映が遅れ、操作推論学習モデル７０は、より大きな反応を求めて、必要以上に大きくアクセルペダルを操作してしまう。

これに対し、本変形例においては、調整部５５によって、仮想装置モデル５０の入出力の分解能を試験装置１にあわせて調整することができる。これにより、上記のような必要以上に大きな操作を抑制し、実車両への負担を低減可能である。

本変形例が、既に説明した実施形態と同様な他の効果を奏することは言うまでもない。

[0071] なお、本発明のドライブ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及び各変形例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。

[0072] 例えば、上記第１変形例においては、調整部５５が推論操作調整部５６と車両モデル入力調整部５７を備えておらず、模擬走行状態調整部５８のみを備え、模擬走行状態調整部５８が、仮想装置モデル５０を、試験装置１に対応する１つの制御処理系と見做して、仮想装置モデル５０全体における時間をまとめて調整していた。

これに変えて、調整部５５が車両モデル入力調整部５７と模擬走行状態調整部５８を備えておらず、推論操作調整部５６のみを備えた構成としてもよい。この場合においては、推論操作調整部５６が、推論操作が推論されてから、推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を調整する。

[0073] あるいは、調整部５５が、推論操作調整部５６、車両モデル入力調整部５７、及び模擬走行状態調整部５８のいずれか２つを備えた構成としてもよい

。この場合においては、推論操作調整部５６、車両モデル入力調整部５７、及び模擬走行状態調整部５８のいずれか２つが、推論操作が推論されてから、推論操作がドライブロボット４に入力されて車両２が操作、走行され、車両２の走行状態が取得され、走行状態が操作推論学習モデル７０に適用されるまでに要する実時間を基に、模擬走行状態ｏが出力されてから操作推論学習モデル７０に適用されるまでの時間を、分担して調整する。

[0074] また、上記実施形態及び各変形例においては、車両モデル５２はニューラルネットワークとして実現された車両学習モデル６０を備え、この車両学習モデル６０によって車両２を模擬動作させていたが、これに限られない。すなわち、車両学習モデルは、ニューラルネットワーク以外の手段によって機械学習された機械学習モデルであって構わない。あるいは、車両モデルは、機械学習された学習モデルを備えた構成でなくともよく、例えば数式モデル等で実現されていてもよい。

このようにした場合においては、何らかの車両モデルが用意できる環境にあるのであれば、車両学習モデル６０を機械学習させなくとも、操作推論学習モデル７０を事前学習することができる。したがって、操作推論学習モデル７０の学習が容易である。

[0075] これ以外にも、本発明の主旨を逸脱しない限り、上記実施形態及び各変形例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

[0076] １　試験装置（実環境）

２　車両

２ｃ　アクセルペダル

２ｄ　ブレーキペダル

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　学習システム

１１　学習制御装置

２０　ドライブロボット制御部

３０　学習部

４０　強化学習部

４１　操作内容推論部

４２　状態行動価値推論部

４３　報酬計算部

５０　仮想装置モデル

５１　ドライブロボットモデル（自動操縦ロボットモデル）

５２　車両モデル

５３　シャシーダイナモメータモデル

５５、５５Ａ　調整部

５６　推論操作調整部

５７　車両モデル入力調整部

５８、５８Ａ　模擬走行状態調整部

６０　車両学習モデル

７０　操作推論学習モデル

８０　価値推論学習モデル

ｉ１　車速系列

ｉ２　アクセルペダル操作量系列（入力操作）

ｉ３　ブレーキペダル操作量系列（入力操作）

ｏ　模擬走行状態

請求の範囲

[請求項1] 車両と、前記車両に搭載された自動操縦ロボットとを備える実環境と、車速を含む前記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を備え、前記自動操縦ロボットは、前記操作推論学習モデルが推論した推論操作を基に当該車両を走行させ、前記操作推論学習モデルを機械学習する、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、

前記車両を模擬動作するように設定され、前記推論操作を基に、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、車両モデルを備えた、仮想装置モデルを備え、

前記仮想装置モデルは、前記推論操作が入力されると、当該推論操作を基に、前記車両モデルにより前記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機械学習し、

前記実環境において、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボットに入力されて前記車両が操作、走行され、前記車両の前記走行状態が取得され、当該走行状態が前記操作推論学習モデルに適用されるまでに要する実時間を基に、前記推論操作が推論されてから前記仮想装置モデルにおいて使用され、前記模擬走行状態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、調整部を備えている、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] 前記仮想装置モデルは、前記自動操縦ロボットを模擬動作するように設定され、前記推論操作を入力として、前記車両モデルへ入力される入力操作を出力する、自動操縦ロボットモデルを更に備え、

前記調整部は、

前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボ

ットに入力されるまでの時間を基に、前記推論操作が推論されてから前記自動操縦ロボットモデルに入力されるまでの時間を調整する、推論操作調整部と、

前記推論操作が前記自動操縦ロボットに入力されてから前記車両が操作されるまでの時間を基に、前記入力操作が前記車両モデルに入力されるまでの時間を調整する、車両モデル入力調整部と、

前記車両の前記走行状態が取得されてから前記操作推論学習モデルに適用されるまでの時間を基に、前記模擬走行状態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、模擬走行状態調整部と、

を備えている、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] 前記調整部は、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボットに入力されて前記車両が操作、走行され、前記車両の前記走行状態が取得され、当該走行状態が前記操作推論学習モデルに適用されるまでに要する前記実時間を基に、前記模擬走行状態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、模擬走行状態調整部を備えている、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4] 前記調整部は、前記推論操作の分解能を、前記仮想装置モデルにおける分解能へと変換し、前記模擬走行状態の分解能を、前記車両に対して取得される前記走行状態の分解能へと変換する、請求項１から３のいずれか一項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5] 前記操作は、アクセルペダルとブレーキペダルのいずれか一方または双方のペダルの、ペダル操作量を含む、請求項１から４のいずれか一項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6] 前記車両モデルは、前記車両を模擬動作するように機械学習され、前記推論操作を基に、前記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は、ニューラルネットワークで実現されている、請求項１から５のいずれか一項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7] 前記操作推論学習モデルは、強化学習されている、請求項１から６のいずれか一項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8] 車両と、前記車両に搭載された自動操縦ロボットとを備える実環境と、車速を含む前記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に関し、前記自動操縦ロボットは、前記操作推論学習モデルが推論した推論操作を基に当該車両を走行させ、前記操作推論学習モデルを機械学習する、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、

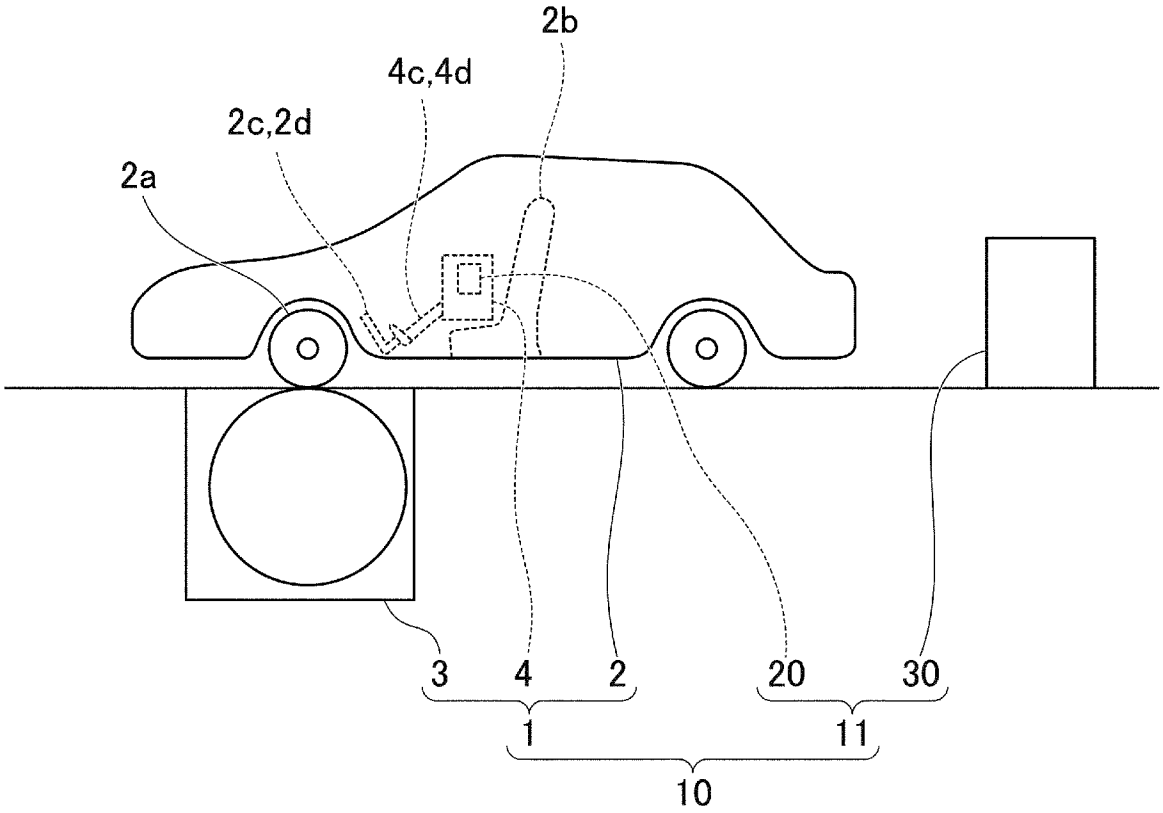
前記車両を模擬動作するように設定され、前記推論操作を基に、前記車両を模した前記走行状態である模擬走行状態を出力する、車両モデルを備えた、仮想装置モデルにおいて、前記推論操作が入力されると、当該推論操作を基に、前記車両モデルにより前記模擬走行状態を出力し、

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機械学習し、

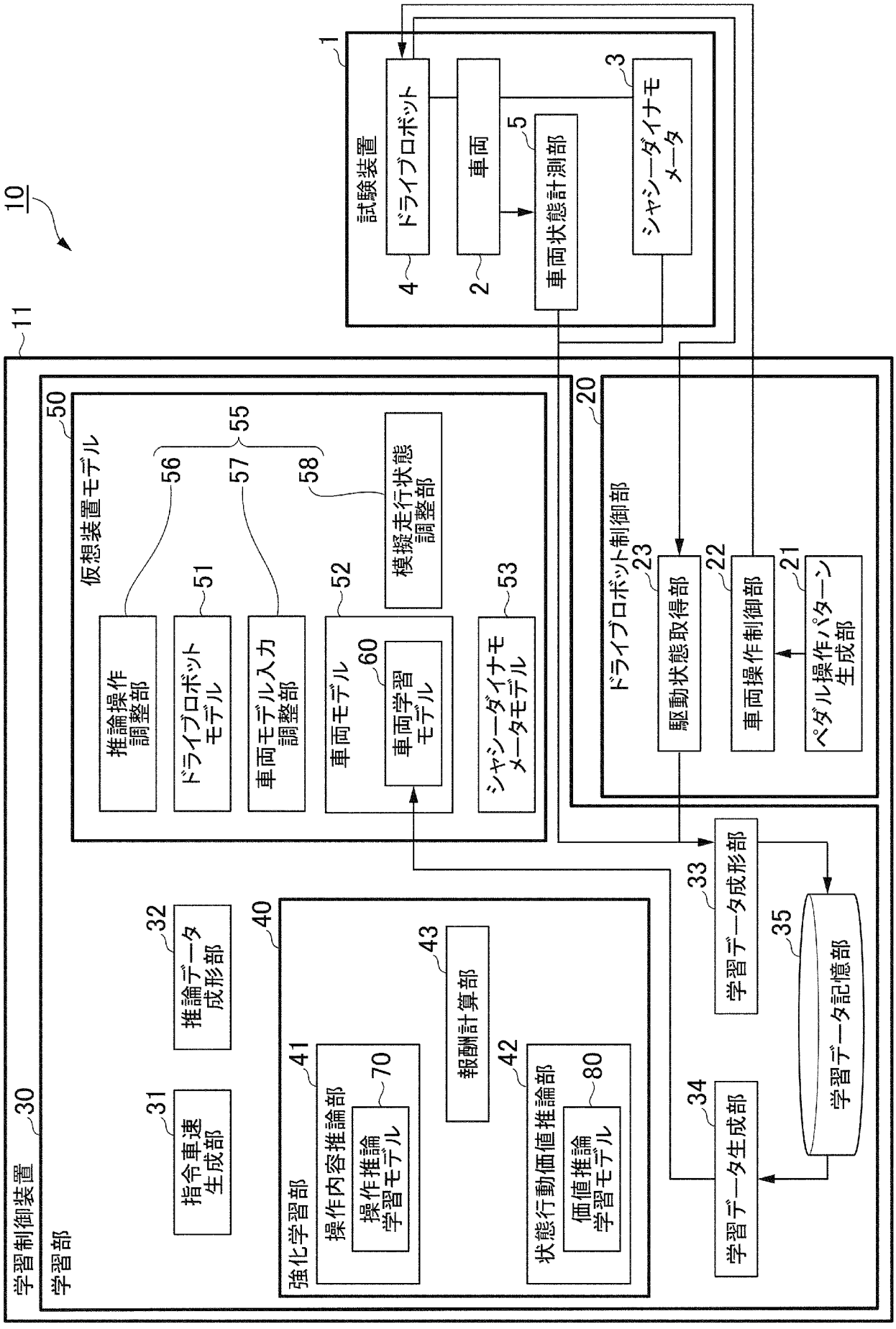
前記実環境において、前記推論操作が推論されてから、当該推論操作が前記自動操縦ロボットに入力されて前記車両が操作、走行され、前記車両の前記走行状態が取得され、当該走行状態が前記操作推論学習モデルに適用されるまでに要する実時間を基に、前記推論操作が推論されてから前記仮想装置モデルにおいて使用され、前記模擬走行状

態が出力されてから前記操作推論学習モデルに適用されるまでの時間を調整する、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。

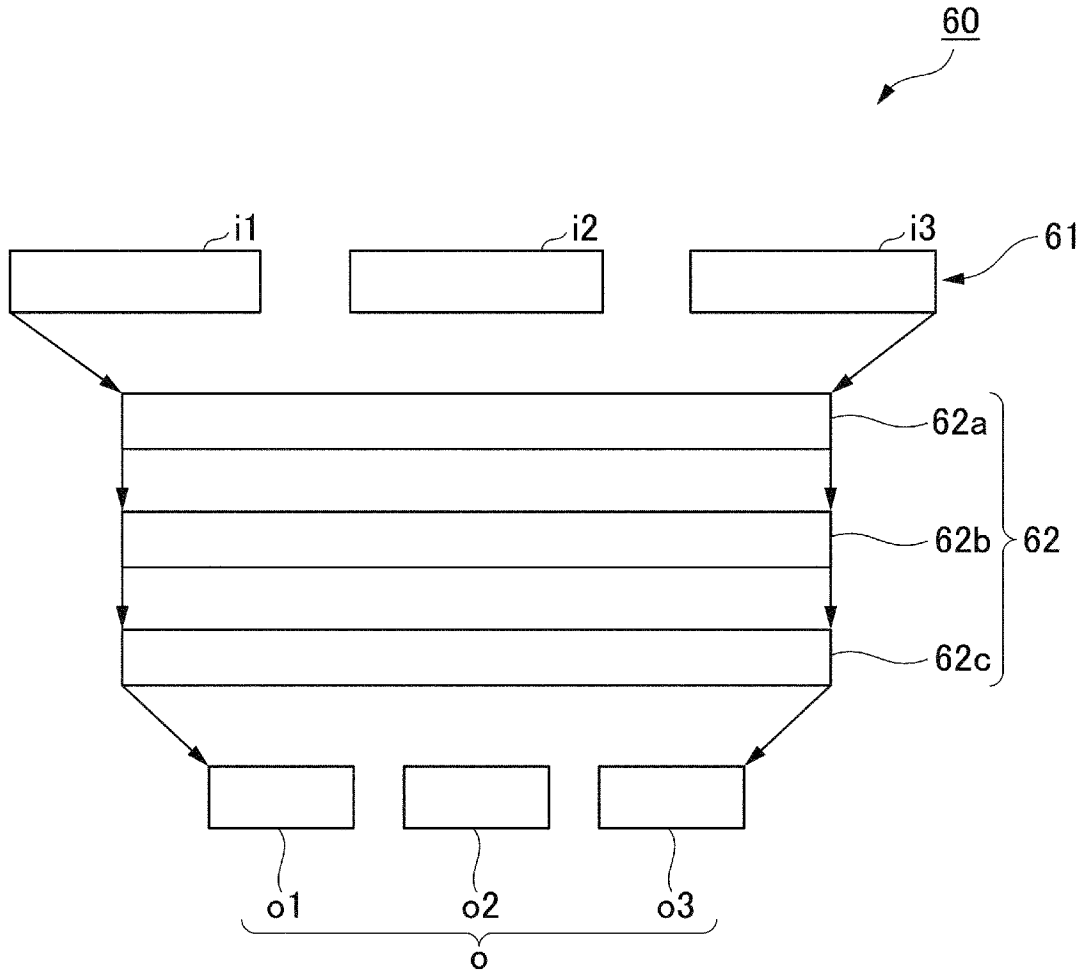
[図1]



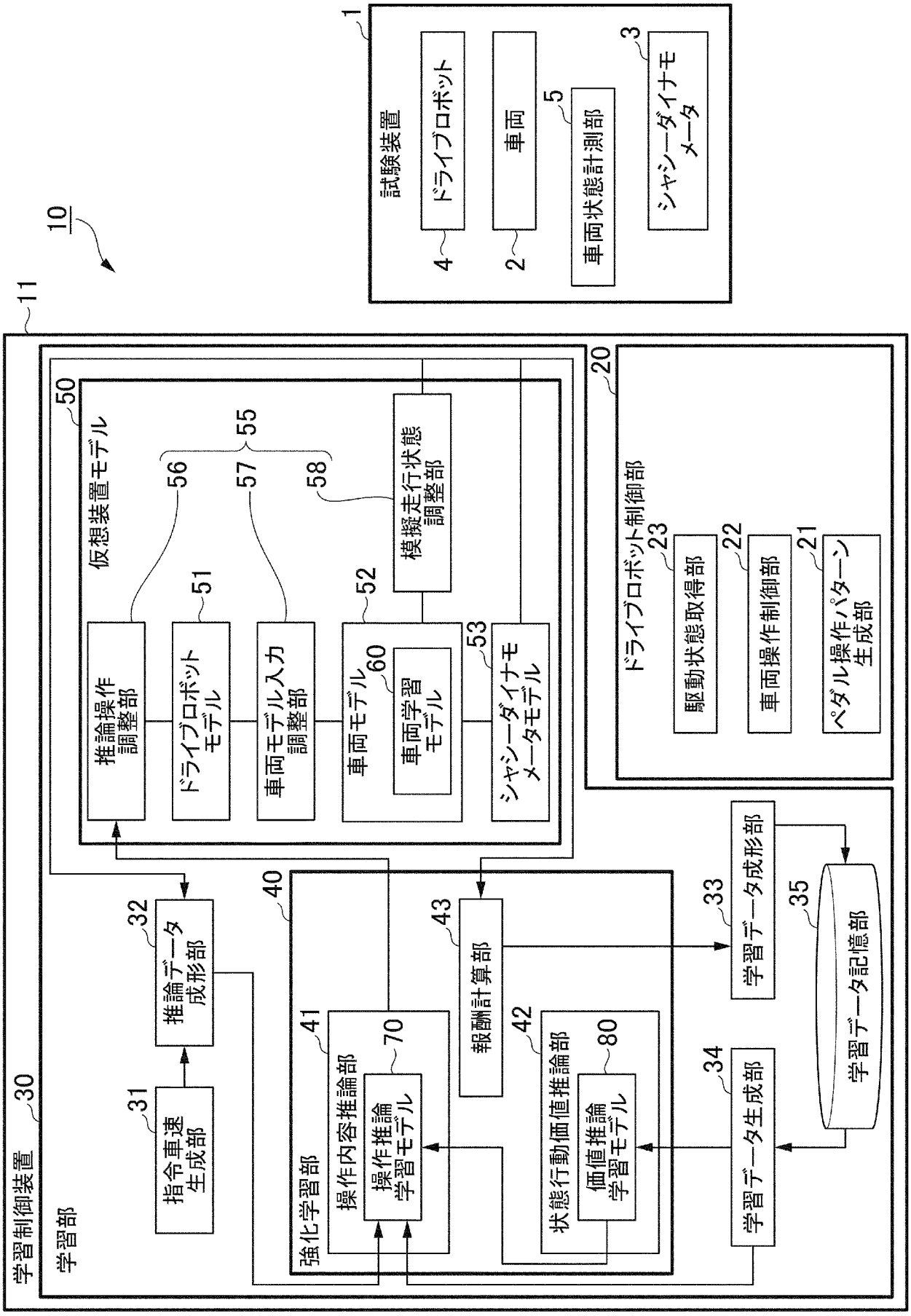
[図2]


10
11
学習制御装置
30
学習部
試験装置
1
ドライブロボット
4
車両
2
車両状態計測部
5
シャシーダイナモメータ
3
50
仮想装置モデル
推論操作調整部
56
ドライブロボットモデル
51
車両モデル入力調整部
57
車両モデル
52
車両学習モデル
60
シャシーダイナモメータモデル
53
模擬走行状態調整部
58
55
20
ドライブロボット制御部
駆動状態取得部
23
車両操作制御部
22
ペダル操作パターン生成部
21
推論データ成形部
32
指令車速生成部
31
40
強化学習部
操作内容推論部
41
操作推論学習モデル
70
報酬計算部
43
状態行動価値推論部
42
価値推論学習モデル
80
学習データ成形部
33
学習データ記憶部
35
学習データ生成部
34

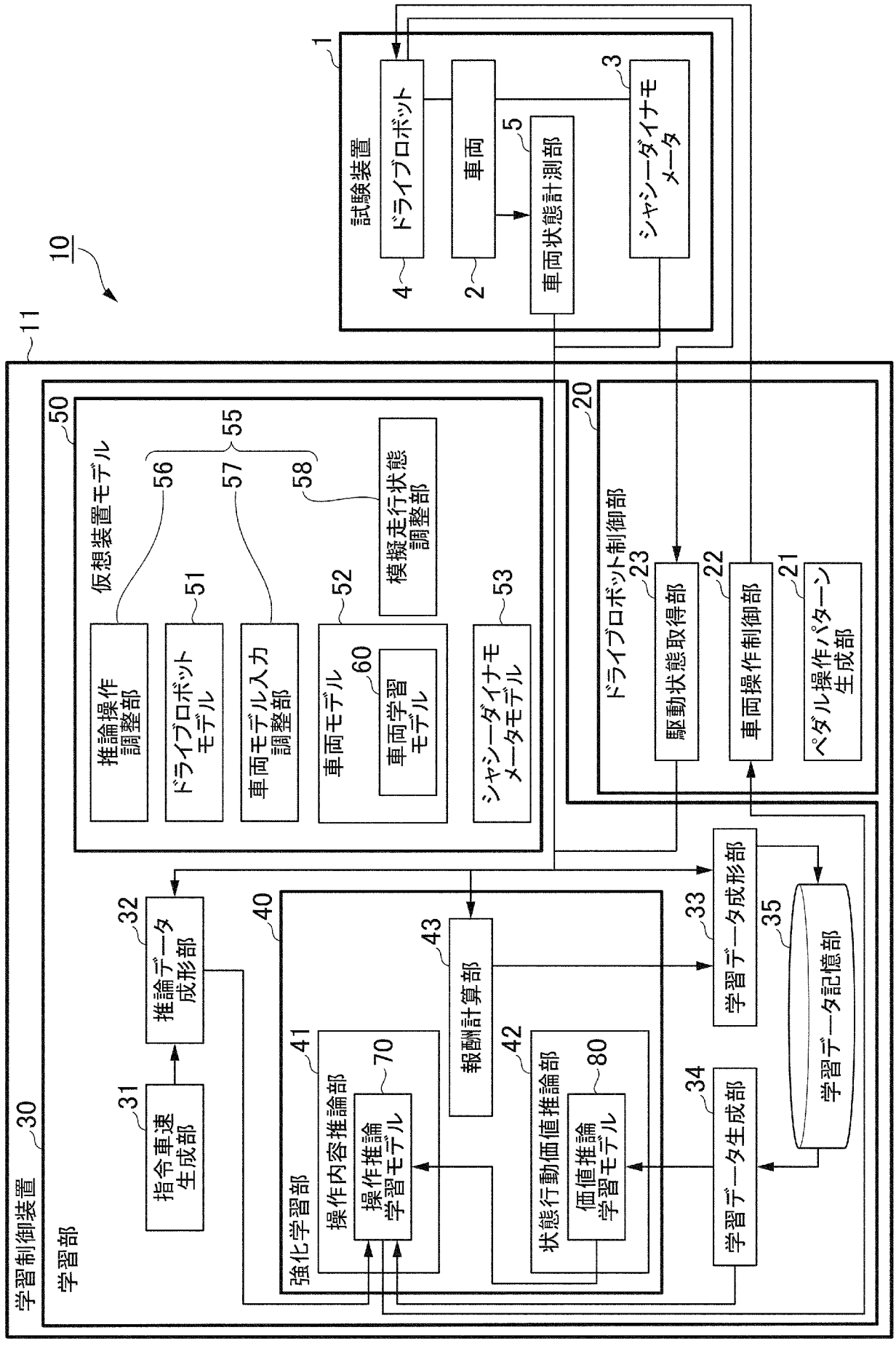
[図3]



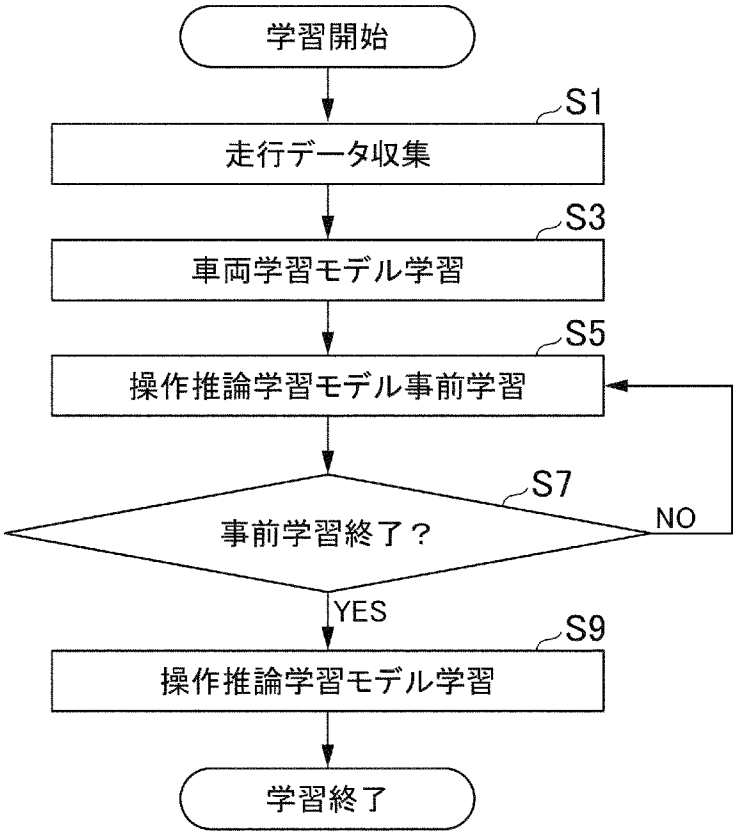
[図4]


10
1
試験装置
ドライブロボット
4
車両
2
車両状態計測部
5
シャシーダイナモメータ
3
11
学習制御装置
30
学習部
50
仮想装置モデル
55
56
57
58
推論操作調整部
ドライブロボットモデル
51
車両モデル入力調整部
車両モデル
52
60
車両学習モデル
模擬走行状態調整部
シャシーダイナモメータモデル
53
20
ドライブロボット制御部
駆動状態取得部
23
車両操作制御部
22
ペダル操作パターン生成部
21
40
強化学習部
操作内容推論部
41
操作推論学習モデル
70
報酬計算部
43
状態行動価値推論部
42
価値推論学習モデル
80
推論データ成形部
32
指令車速生成部
31
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

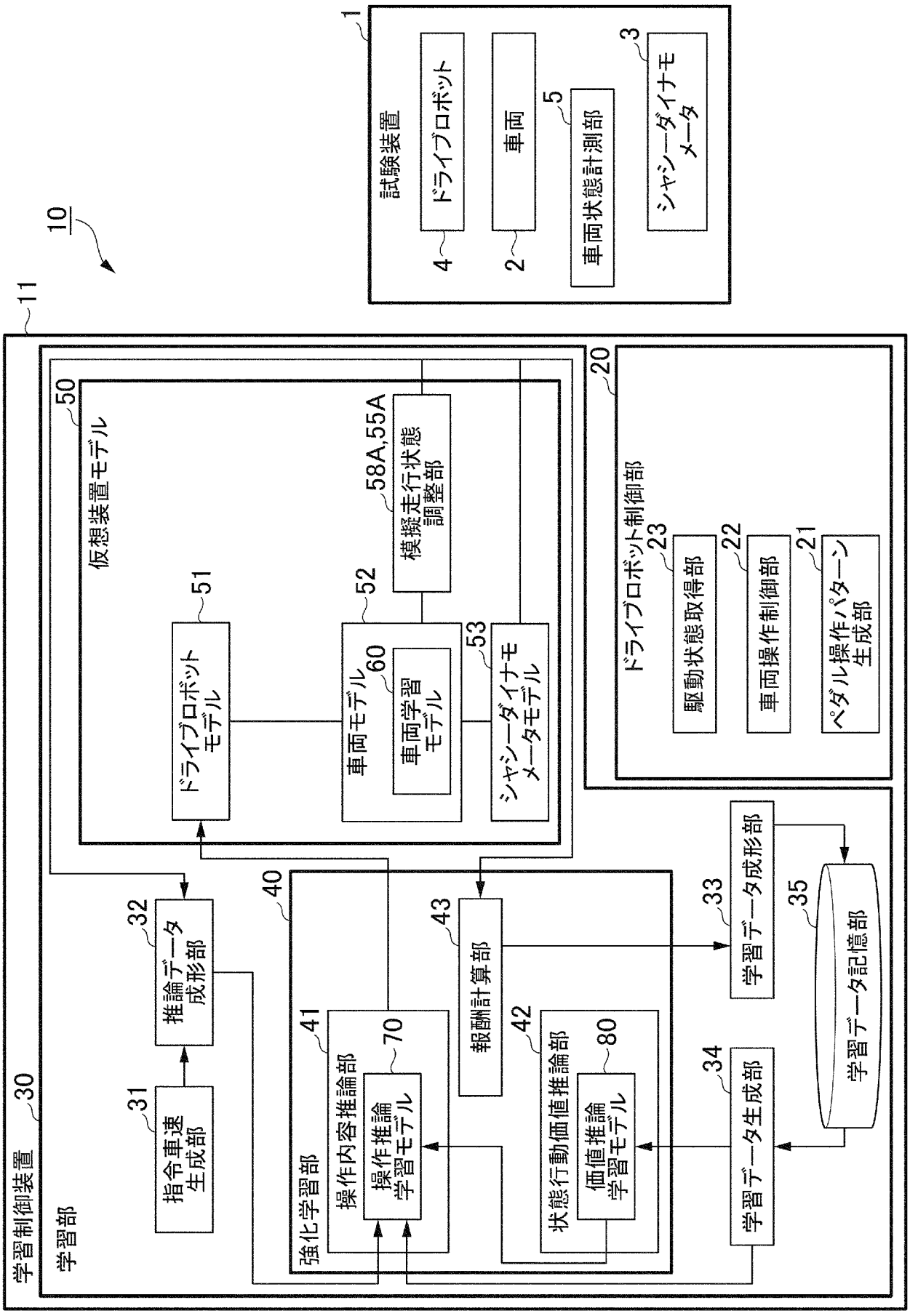
[図5]


10
11
試験装置
1
ドライブロボット
4
車両
2
車両状態計測部
5
シャシーダイナモメータ
3
学習制御装置
30
学習部
仮想装置モデル
50
推論操作調整部
56
ドライブロボットモデル
51
車両モデル入力調整部
57
車両モデル
52
車両学習モデル
60
シャシーダイナモメータモデル
53
55
模擬走行状態調整部
58
ドライブロボット制御部
20
駆動状態取得部
23
車両操作制御部
22
ペダル操作パターン生成部
21
指令車速生成部
31
推論データ成形部
32
強化学習部
40
操作内容推論部
41
操作推論学習モデル
70
報酬計算部
43
状態行動価値推論部
42
価値推論学習モデル
80
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

[図6]


学習開始
S1
走行データ収集
S3
車両学習モデル学習
S5
操作推論学習モデル事前学習
S7
事前学習終了？
NO
YES
S9
操作推論学習モデル学習
学習終了

[図7]


10
1
試験装置
4
ドライブロボット
2
車両
5
車両状態計測部
3
シャシーダイナモ
メータ
11
学習制御装置
30
学習部
50
仮想装置モデル
51
ドライブロボット
モデル
52
車両モデル
60
車両学習
モデル
58A,55A
模擬走行状態
調整部
53
シャシーダイナモ
メータモデル
20
ドライブロボット制御部
23
駆動状態取得部
22
車両操作制御部
21
ペダル操作パターン
生成部
32
推論データ
成形部
31
指令車速
生成部
40
強化学習部
41
操作内容推論部
70
操作推論
学習モデル
43
報酬計算部
42
状態行動価値推論部
80
価値推論
学習モデル
33
学習データ成形部
35
学習データ記憶部
34
学習データ生成部

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2021/032055

A. CLASSIFICATION OF SUBJECT MATTER

F02D 29/02(2006.01)i; ***G01M 17/007***(2006.01)i
FI: G01M17/007 B; F02D29/02 E

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)

F02D29/00-29/06; G01M17/00-17/06

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 1971-2021
Registered utility model specifications of Japan 1996-2021
Published registered utility model applications of Japan 1994-2021

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 2020-056737 A (MEIDENSHA ELECTRIC MFG CO LTD) 09 April 2020 (2020-04-09) paragraphs [0011]-[0016], [0086]-[0090], [0107], fig. 1-2, 7	1-8
A	CN 110795818 A (TENCENT TECH SHENZHEN CO., LTD.) 14 February 2020 (2020-02-14) paragraphs [0055]-[0083], fig. 1-4	1-8
A	US 2019/0278272 A1 (SZ DJI TECHNOLOGY CO., LTD.) 12 September 2019 (2019-09-12) paragraphs [0041]-[0076], fig. 1-3	1-8
A	US 2006/0155734 A1 (GRIMES, Michael R.) 13 July 2006 (2006-07-13) paragraphs [0003], [0014]-[0021], fig. 1	1-8

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 October 2021	**09 November 2021**
Name and mailing address of the ISA/JP **Japan Patent Office (ISA/JP)** **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/JP2021/032055

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2020-056737 A	09 April 2020	(Family: none)	
CN 110795818 A	14 February 2020	(Family: none)	
US 2019/0278272 A1	12 September 2019	WO 2018/098658 A1	
		CN 107004039 A	
US 2006/0155734 A1	13 July 2006	DE 102006000916 B3	
		CN 1800809 A	

国際調査報告

国際出願番号
PCT/JP2021/032055

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

F02D 29/02(2006.01)i; G01M 17/007(2006.01)i
FI: G01M17/007 B; F02D29/02 E

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））

F02D29/00-29/06; G01M17/00-17/06

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	１９７１－２０２１年
日本国実用新案登録公報	１９９６－２０２１年
日本国登録実用新案公報	１９９４－２０２１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 2020-056737 A (株式会社明電舎) 09.04.2020 (2020-04-09) 段落[0011]-[0016],[0086]-[0090],[0107], 図1-2,7	1-8
A	CN 110795818 A (TENCENT TECH SHENZHEN CO., LTD.) 14.02.2020 (2020-02-14) 段落[0055]-[0083], 図1-4	1-8
A	US 2019/0278272 A1 (SZ DJI TECHNOLOGY CO., LTD.) 12.09.2019 (2019-09-12) 段落[0041]-[0076], 図1-3	1-8
A	US 2006/0155734 A1 (GRIMES, Michael R.) 13.07.2006 (2006-07-13) 段落[0003], [0014]-[0021], 図1	1-8

☐ 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
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
27.10.2021	09.11.2021

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）
瓦井　秀憲 2J 4645
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/JP2021/032055

引用文献	公表日	パテントファミリー文献	公表日
JP 2020-056737 A	09.04.2020	（ファミリーなし）	
CN 110795818 A	14.02.2020	（ファミリーなし）	
US 2019/0278272 A1	12.09.2019	WO 2018/098658 A1	
		CN 107004039 A	
US 2006/0155734 A1	13.07.2006	DE 102006000916 B3	
		CN 1800809 A	