

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第6744597号  
(P6744597)

(45) 発行日 令和2年8月19日(2020.8.19)

(24) 登録日 令和2年8月4日(2020.8.4)

(51) Int.Cl.	F I				
<b>FO2D 45/00</b>	<b>(2006.01)</b>	FO2D 45/00	376		
<b>FO2D 43/00</b>	<b>(2006.01)</b>	FO2D 43/00	310Z		
<b>B6OW 30/182</b>	<b>(2020.01)</b>	B6OW 30/182			
<b>G05B 11/36</b>	<b>(2006.01)</b>	G05B 11/36		G	
<b>G16Y 10/40</b>	<b>(2020.01)</b>	G16Y 10/40			

請求項の数 14 (全 24 頁) 最終頁に続く

(21) 出願番号 特願2019-191093 (P2019-191093)  
 (22) 出願日 令和1年10月18日(2019.10.18)  
 審査請求日 令和1年11月20日(2019.11.20)

早期審査対象出願

(73) 特許権者 000003207  
 トヨタ自動車株式会社  
 愛知県豊田市トヨタ町1番地  
 (74) 代理人 100105957  
 弁理士 恩田 誠  
 (74) 代理人 100068755  
 弁理士 恩田 博宣  
 (72) 発明者 橋本 洋介  
 愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内  
 (72) 発明者 片山 章弘  
 愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内

最終頁に続く

(54) 【発明の名称】 車両用制御データの生成方法、車両用制御装置、車両用制御システム、および車両用学習装置

(57) 【特許請求の範囲】

【請求項1】

複数の走行制御モードのうちの1つを選択可能な車両に適用され、  
 前記車両の状態と前記車両内の電子機器の操作に関する変数である行動変数との関係を規定する関係規定データが記憶装置に記憶された状態で、  
 前記電子機器を操作する操作処理と、  
 前記車両の状態を検出するセンサの検出値を取得する取得処理と、  
 前記取得処理によって取得された前記検出値に基づき、前記走行制御モードと相関を有する前記車両の特性が基準を満たす場合に満たさない場合よりも大きい報酬を与える報酬算出処理と、  
 前記取得処理によって取得された前記検出値に基づく前記車両の状態、前記電子機器の操作に用いられた行動変数の値、および該操作に対して与えられた前記報酬を予め定められた更新写像への入力とし、前記関係規定データを更新する更新処理と、  
 を実行装置に実行させ、  
 前記更新写像は、前記関係規定データに従って前記電子機器が操作される場合の前記報酬についての期待収益を増加させるように更新された前記関係規定データを出力するものであり、  
 前記報酬算出処理は、前記選択される走行制御モードが互いに異なる第1の走行制御モードと第2の走行制御モードとで、前記特性が同一の基準を満たす場合であっても異なる報酬を与える処理を含む車両用制御データの生成方法。

## 【請求項 2】

複数の走行制御モードのうちの1つを選択可能な車両に適用され、  
前記車両の状態と前記車両内の電子機器の操作に関する変数である行動変数との関係を規定する関係規定データが記憶装置に記憶された状態で、

前記電子機器を操作する操作処理と、

前記車両の状態を検出するセンサの検出値を取得する取得処理と、

前記取得処理によって取得された前記検出値に基づき、前記走行制御モードと相関を有する前記車両の特性が基準を満たす場合に満たさない場合よりも大きい報酬を与える報酬算出処理と、

前記取得処理によって取得された前記検出値に基づく前記車両の状態、前記電子機器の操作に用いられた行動変数の値、および該操作により生じる前記車両の特性に対して与えられた前記報酬を予め定められた更新写像への入力とし、前記関係規定データを更新する更新処理と、

を実行装置に実行させ、

前記更新写像は、前記関係規定データに従って前記電子機器が操作される場合の前記報酬についての期待収益を増加させるように更新された前記関係規定データを出力するものであり、

前記報酬算出処理は、前記選択される走行制御モードが互いに異なる第1の走行制御モードと第2の走行制御モードとで、前記特性が同一の基準を満たす場合であっても異なる報酬を与える処理を含む車両用制御データの生成方法。

## 【請求項 3】

前記車両の状態には、前記車両に搭載される推力生成装置のトルクが含まれ、

前記報酬算出処理は、前記トルクとトルク指令値の差の絶対値が小さい場合に大きい場合よりもより大きい報酬を与えて且つ、前記複数の走行制御モードのうちの第1の走行制御モードと第2の走行制御モードとで前記絶対値が同一であっても互いに異なる報酬を与える処理を含む請求項1または2記載の車両用制御データの生成方法。

## 【請求項 4】

前記車両の状態には、前記車両に搭載される変速装置の状態が含まれ、

前記報酬算出処理は、前記変速装置の状態が基準を満たす場合に満たさない場合よりも大きい報酬を与えて且つ、前記複数の走行制御モードのうちの第1の走行制御モードと第2の走行制御モードとで前記変速装置の状態が同一の基準を満たす場合であっても互いに異なる報酬を与える処理を含み、

前記変速装置の状態が基準を満たす場合に満たさない場合よりも大きい報酬を与える処理は、前記変速装置による変速比の切り替えに要する時間が所定時間以内である場合に所定時間を超える場合よりも大きい報酬を与える処理と、前記変速装置の入力軸52の回転速度の変化速度の絶対値が入力側所定値以下である場合に入力側所定値を超える場合よりも大きい報酬を与える処理と、前記変速装置の出力軸54の回転速度の変化速度の絶対値が出力側所定値以下である場合に入力側所定値を超える場合よりも大きい報酬を与える処理との、3つの処理のうちの少なくとも1つの処理を含む請求項1または2記載の車両用制御データの生成方法。

## 【請求項 5】

前記車両の状態には、前記車両におけるエネルギー消費率が含まれ、

前記報酬算出処理は、前記エネルギー消費率が低い場合に高い場合よりも大きい報酬を与えて且つ、前記複数の走行制御モードのうちの第1の走行制御モードと第2の走行制御モードとで前記エネルギー消費率が同一の基準を満たす場合であっても互いに異なる報酬を与える処理を含む請求項1または2記載の車両用制御データの生成方法。

## 【請求項 6】

前記車両の状態には、アクセル操作量の変化が含まれ、

前記報酬算出処理は、前記アクセル操作量の変化に伴う前記車両の前後方向の加速度が基準を満たす場合に満たさない場合よりも大きい報酬を与えて且つ、前記複数の走行制御

10

20

30

40

50

モードのうちの第1の走行制御モードと第2の走行制御モードとで前記加速度が同一の基準を満たす場合であっても互いに異なる報酬を与える処理を含む請求項1または2記載の車両用制御データの生成方法。

【請求項7】

前記車両は、当該車両の推力生成装置として内燃機関を備えるものであり、  
前記電子機器には、前記内燃機関のスロットルバルブが含まれ、  
前記行動変数には、前記スロットルバルブの開口度に関する変数が含まれる請求項6記載の車両用制御データの生成方法。

【請求項8】

前記電子機器には、前記内燃機関の点火装置が含まれ、  
前記行動変数には、点火時期に関する変数が含まれる請求項7記載の車両用制御データの生成方法。

10

【請求項9】

前記関係規定データは、前記アクセル操作量の時系列データと前記行動変数との関係を規定するデータである請求項6～8のいずれか1項に記載の車両用制御データの生成方法。

【請求項10】

前記更新処理によって更新された前記関係規定データに基づき、前記車両の状態と前記期待収益を最大化する前記行動変数の値とを1対1に対応付けることによって前記車両の状態を入力とし前記期待収益を最大化する前記行動変数の値を出力する制御用写像データを生成する処理を前記実行装置に実行させる請求項1～9のいずれか1項に記載の車両用制御データの生成方法。

20

【請求項11】

請求項1～9のいずれか1項に記載の前記記憶装置および前記実行装置を備え、  
前記操作処理は、前記関係規定データに基づき前記車両の状態に応じた行動変数の値に従って前記電子機器を操作する処理を含み、  
前記取得処理は、前記複数の走行制御モードのうちのいずれをユーザが選択したかの選択結果を取得する処理を含む車両用制御装置。

【請求項12】

請求項11記載の前記実行装置および前記記憶装置を備え、  
前記実行装置は、前記車両に搭載される第1実行装置と、車載装置とは別の第2実行装置と、を含み、  
前記第1実行装置は、少なくとも前記取得処理および前記操作処理を実行し、  
前記第2実行装置は、少なくとも前記更新処理を実行する車両用制御システム。

30

【請求項13】

請求項12記載の第1実行装置を備える車両用制御装置。

【請求項14】

請求項12記載の第2実行装置を備える車両用学習装置。

【発明の詳細な説明】

【技術分野】

40

【0001】

本発明は、車両用制御データの生成方法、車両用制御装置、車両用制御システム、および車両用学習装置に関する。

【背景技術】

【0002】

たとえば下記特許文献1には、アクセルペダルの操作量をフィルタ処理した値に基づき、車両に搭載される内燃機関の操作部としてのスロットルバルブを操作する制御装置が記載されている。

【先行技術文献】

【特許文献】

50

## 【 0 0 0 3 】

【特許文献 1】特開 2 0 1 6 - 6 3 2 7 号公報

## 【発明の概要】

【発明が解決しようとする課題】

## 【 0 0 0 4 】

ところで、上記フィルタは、アクセルペダルの操作量に応じて車両に搭載される内燃機関のスロットルバルブの操作量を適切な操作量に設定するものである必要があることから、その適合には熟練者が多くの工数をかける必要が生じる。このように、従来は、車両の状態に応じた車両内の電子機器の操作量等の適合には、熟練者が多くの工数をかけていた。

10

【課題を解決するための手段】

## 【 0 0 0 5 】

以下、上記課題を解決するための手段およびその作用効果について記載する。

1. 複数の走行制御モードのうちの 1 つを選択可能な車両に適用され、前記車両の状態と前記車両内の電子機器の操作に関する変数である行動変数との関係を規定する関係規定データが記憶装置に記憶された状態で、前記電子機器を操作する操作処理と、前記車両の状態を検出するセンサの検出値を取得する取得処理と、前記取得処理によって取得された前記検出値に基づき、前記走行制御モードと相関を有する前記車両の特性が基準を満たす場合に満たさない場合よりも大きい報酬を与える報酬算出処理と、前記取得処理によって取得された前記検出値に基づく前記車両の状態、前記電子機器の操作に用いられた行動変数の値、および該操作に対応する前記報酬を予め定められた更新写像への入力とし、前記関係規定データを更新する更新処理と、を実行装置に実行させ、前記更新写像は、前記関係規定データに従って前記電子機器が操作される場合の前記報酬についての期待収益を増加させるように更新された前記関係規定データを出力するものであり、前記報酬算出処理は、前記選択される走行制御モードが互いに異なる第 1 の走行制御モードと第 2 の走行制御モードとで、前記特性が同一の基準を満たす場合であっても異なる報酬を与える処理を含む車両用制御データの生成方法である。

20

## 【 0 0 0 6 】

上記方法では、電子機器の操作に伴う報酬を算出することによって、当該操作によってどのような報酬が得られるかを把握することができる。そして、報酬に基づき、強化学習に従った更新写像によって関係規定データを更新することにより、車両の状態と行動変数との関係を設定することができる。したがって、車両の状態と行動変数との関係の設定に際して、熟練者に要求される工数を削減できる。

30

## 【 0 0 0 7 】

ところで、複数の走行制御モードのうちの 1 つを選択可能な車両においては、走行制御モードごとに、車両の状態にとって適切な行動変数の値が異なりうる。そこで上記方法では、第 1 の走行制御モードと第 2 の走行制御モードとで車両の特性が同一の基準を満たす場合であっても異なる報酬を与えることにより、走行制御モードごとに、車両の状態にとって適切な行動変数の値を強化学習によって学習することが可能となる。

## 【 0 0 0 8 】

2. 前記車両の状態には、アクセル操作量の変化が含まれ、前記報酬算出処理は、前記アクセル操作量の変化に伴う前記車両の前後方向の加速度が基準を満たす場合に満たさない場合よりも大きい報酬を与えて且つ、前記複数の走行制御モードのうちの第 1 の走行制御モードと第 2 の走行制御モードとで前記加速度が同一の基準を満たす場合であっても互いに異なる報酬を与える処理を含む上記 1 記載の車両用制御データの生成方法である。

40

## 【 0 0 0 9 】

アクセル操作量の変化によって生じる車両の前後方向の加速度の大きさは、車両の走行制御モードに大いに関係することから、上記方法のように、走行制御モードに応じて加速度が同一の基準を満たす場合であっても互いに異なる報酬を与えることにより、走行制御モードごとに、アクセル操作にとって適切な行動変数の値を強化学習によって学習できる

50

。

【0010】

3. 前記車両は、当該車両の推力生成装置として内燃機関を備えるものであり、前記電子機器には、前記内燃機関のスロットルバルブが含まれ、前記行動変数には、前記スロットルバルブの開口度に関する変数が含まれる上記2記載の車両用制御データの生成方法である。

【0011】

たとえば吸入空気量に応じて噴射量を調整する内燃機関等においては、スロットルバルブの開口度に応じて内燃機関のトルクや出力が大きく変化する。そのため、アクセル操作量に対する行動変数としてスロットルバルブの開口度に関する変数を用いることにより、車両の推進力を好適に調整することができる。

10

【0012】

4. 前記電子機器には、前記内燃機関の点火装置が含まれ、前記行動変数には、点火時期に関する変数が含まれる上記3記載の車両用制御データの生成方法である。

点火装置を備える内燃機関においては、点火装置によって操作される点火時期によって、内燃機関のトルクを調整できる。そのため、上記方法では、行動変数に点火時期に関する変数を含めることにより、アクセル操作量の変化に応じて車両の推進力を好適に調整できる。

【0013】

5. 前記関係規定データは、前記アクセル操作量の時系列データと前記行動変数との関係を規定するデータである上記2～4のいずれか1つに記載の車両用制御データの生成方法である。

20

【0014】

上記方法では、アクセル操作量の時系列データと行動変数との関係を規定することにより、たとえばアクセル操作量の変化量を示す1次元のデータと行動変数との関係を規定する場合と比較して、アクセル操作量の様々な変化に対して行動変数の値をきめ細かく調整できる。

【0015】

6. 前記更新処理によって更新された前記関係規定データに基づき、前記車両の状態と前記期待収益を最大化する前記行動変数の値とを1対1に対応付けることによって前記車両の状態を入力とし前記期待収益を最大化する前記行動変数の値を出力する制御用写像データを生成する処理を前記実行装置に実行させる上記1～5のいずれか1つに記載の車両用制御データの生成方法である。

30

【0016】

上記方法では、強化学習によって学習された関係規定データに基づき、制御用写像データを生成する。そのため、その制御用写像データを制御装置に実装することにより、車両の状態に基づき、期待収益を最大化する行動変数の値を簡易に設定することが可能となる。

。

【0017】

7. 上記1～5のいずれか1つに記載の前記記憶装置および前記実行装置を備え、前記操作処理は、前記関係規定データに基づき前記車両の状態に応じた行動変数の値に従って前記電子機器を操作する処理を含み、前記取得処理は、前記複数の走行制御モードのうちのいずれをユーザが選択したかの選択結果を取得する処理を含む車両用制御装置である。

40

【0018】

上記構成では、強化学習によって学習された関係規定データに基づき行動変数の値が設定され、それに基づき電子機器が操作されることにより、期待収益を大きくするように電子機器を操作することができる。しかも、ユーザによる走行制御モードの選択結果を取得することにより、車両の走行に伴って走行制御モードごとに、関係規定データを更新できる。

【0019】

50

8. 上記7記載の前記実行装置および前記記憶装置を備え、前記実行装置は、前記車両に搭載される第1実行装置と、車載装置とは別の第2実行装置と、を含み、前記第1実行装置は、少なくとも前記取得処理および前記操作処理を実行し、前記第2実行装置は、少なくとも前記更新処理を実行する車両用制御システムである。

【0020】

上記構成では、更新処理を第2実行装置によって実行することにより、更新処理を第1実行装置が実行する場合と比較して、第1実行装置の演算負荷を軽減できる。

なお、第2実行装置が車載装置とは別の装置であることは、第2実行装置が車載装置ではないことを意味する。

【0021】

9. 上記8記載の第1実行装置を備える車両用制御装置である。

10. 上記8記載の第2実行装置を備える車両用学習装置である。

【図面の簡単な説明】

【0022】

【図1】第1の実施形態にかかる制御装置および駆動系を示す図。

【図2】同実施形態にかかる制御装置が実行する処理の手順を示す流れ図。

【図3】同実施形態にかかるマップデータを生成するシステムを示す図。

【図4】同実施形態にかかる学習処理の手順を示す流れ図。

【図5】同実施形態にかかる学習処理の一部の詳細を示す流れ図。

【図6】同実施形態にかかるマップデータの生成処理の手順を示す流れ図。

【図7】第2の実施形態にかかる制御装置および駆動系を示す図。

【図8】同実施形態にかかる制御装置が実行する処理の手順を示す流れ図。

【図9】第3の実施形態にかかるシステムの構成を示す図。

【図10】(a)および(b)は、同実施形態にかかるシステムが実行する処理の手順を示す流れ図。

【発明を実施するための形態】

【0023】

以下、車両用制御データの生成方法、車両用制御装置、車両用制御システム、および車両用学習装置にかかる実施形態について図面を参照しつつ説明する。

<第1の実施形態>

図1に、本実施形態にかかる車両VC1の駆動系および制御装置の構成を示す。

【0024】

図1に示すように、内燃機関10の吸気通路12には、上流側から順にスロットルバルブ14および燃料噴射弁16が設けられており、吸気通路12に吸入された空気や燃料噴射弁16から噴射された燃料は、吸気バルブ18の開弁に伴って、シリンダ20およびピストン22によって区画される燃焼室24に流入する。燃焼室24内において、燃料と空気との混合気は、点火装置26の火花放電に伴って燃焼に供され、燃焼によって生じたエネルギーは、ピストン22を介してクランク軸28の回転エネルギーに変換される。燃焼に供された混合気は、排気バルブ30の開弁に伴って、排気として排気通路32に排出される。排気通路32には、排気を浄化する後処理装置としての触媒34が設けられている。

【0025】

クランク軸28には、ロックアップクラッチ42を備えたトルクコンバータ40を介して、変速装置50の入力軸52が機械的に連結可能とされている。変速装置50は、入力軸52の回転速度と出力軸54の回転速度との比である変速比を可変とする装置である。出力軸54には、駆動輪60が機械的に連結されている。

【0026】

制御装置70は、内燃機関10を制御対象とし、その制御量であるトルクや排気成分比率等を制御すべく、スロットルバルブ14、燃料噴射弁16および点火装置26等の内燃機関10の操作部を操作する。また、制御装置70は、トルクコンバータ40を制御対象とし、ロックアップクラッチ42の係合状態を制御すべくロックアップクラッチ42を操

10

20

30

40

50

作する。また、制御装置70は、変速装置50を制御対象とし、その制御量としての変速比を制御すべく変速装置50を操作する。なお、図1には、スロットルバルブ14、燃料噴射弁16、点火装置26、ロックアップクラッチ42、および変速装置50のそれぞれの操作信号MS1~MS5を記載している。

#### 【0027】

制御装置70は、制御量の制御のために、エアフローメータ80によって検出される吸入空気量Gaや、スロットルセンサ82によって検出されるスロットルバルブ14の開度(スロットル開度TA)、クランク角センサ84の出力信号Scrを参照する。また、制御装置70は、アクセルセンサ88によって検出されるアクセルペダル86の踏み込み量(アクセル操作量PA)や、加速度センサ90によって検出される車両VC1の前後方向の加速度Gx、モード選択スイッチ92の操作によって定まるモード変数MVの値を参照する。ここで、モード選択スイッチ92は、車両VC1のユーザが、車両VC1の走行制御モードに関する3つの選択肢のうちの1つを選択するためのヒューマンマシンインターフェースである。ここで、3つの走行制御モードは、スポーツモード、ノーマルモード、およびコンフォートモードであり、ユーザの運転操作に対する車両VC1の応答性能が互いに異なったモードである。すなわち、スポーツモードの応答性能が最も高く、コンフォートモードの応答性能が最も低い。コンフォートモードは、車両VC1の運転に際して同乗者が体感する加速度を抑制するモードである。

#### 【0028】

制御装置70は、CPU72、ROM74、電気的に書き換え可能な不揮発性メモリ(記憶装置76)、および周辺回路78を備え、それらがローカルネットワーク79を介して通信可能とされている。ここで、周辺回路78は、内部の動作を規定するクロック信号を生成する回路や、電源回路、リセット回路等を含む。

#### 【0029】

ROM74には、制御プログラム74aが記憶されている。一方、記憶装置76には、アクセル操作量PAを入力変数とし、スロットル開度TAの指令値(スロットル開度指令値TA\*)を出力変数とするデータと点火装置26の遅角量aopを出力変数とするデータとからなるマップデータDMが記憶されている。ここで、遅角量aopは、予め定められた基準点火時期に対する遅角量であり、基準点火時期は、MBT点火時期とノック限界点とのうちの遅角側の時期である。MBT点火時期は、最大トルクの得られる点火時期(最大トルク点火時期)である。またノック限界点は、ノック限界の高い高オクタン価燃料の使用時に、想定される最良の条件下で、ノッキングを許容できるレベル以内に収めることのできる点火時期の進角限界値である。

#### 【0030】

なお、マップデータとは、入力変数の離散的な値と、入力変数の値のそれぞれに対応する出力変数の値と、の組データである。本実施形態では、マップデータDMとして、ノーマルモード用マップデータDM1と、スポーツモード用マップデータDM2と、コンフォートモード用マップデータDM3とを備える。

#### 【0031】

図2に、本実施形態にかかる制御装置70が実行する処理の手順を示す。図2に示す処理は、ROM74に記憶された制御プログラム74aをCPU72がたとえば所定周期で繰り返し実行することにより実現される。なお、以下では、先頭に「S」が付与された数字によって各処理のステップ番号を示す。

#### 【0032】

図2に示す一連の処理において、CPU72は、まずアクセル操作量PAの6個のサンプリング値「PA(1), PA(2), ... PA(6)」からなる時系列データを取得する(S10)。ここで、時系列データを構成する各サンプリング値は、互いに異なるタイミングにおいてサンプリングされたものである。本実施形態では、一定のサンプリング周期でサンプリングされる場合の、互いに時系列的に隣り合う6個のサンプリング値によって時系列データを構成する。

## 【0033】

次にCPU72は、モード変数MVを取得する(S12)。そしてCPU72は、取得したモード変数MVの値が、ノーマルモードを示すか否かを判定する(S14)。CPU72は、ノーマルモードであると判定する場合(S14: YES)、スロットルバルブ14や点火装置26の操作に用いるマップデータDMとして、ノーマルモード用マップデータDM1を選択する(S16)。これに対し、CPU72は、ノーマルモードではないと判定する場合(S14: NO)、モード変数MVの値がスポーツモードを示すか否かを判定する(S18)。そして、CPU72は、スポーツモードであると判定する場合(S18: YES)、スロットルバルブ14や点火装置26の操作に用いるマップデータDMとして、スポーツモード用マップデータDM2を選択する(S20)。これに対し、CPU72は、スポーツモードではないと判定する場合(S18: NO)、スロットルバルブ14や点火装置26の操作に用いるマップデータDMとして、コンフォートモード用マップデータDM3を選択する(S22)。

10

## 【0034】

CPU72は、S16, S20, S22の処理が完了する場合、マップデータDMのうち選択したデータを用いてスロットル開口度指令値TA\*および遅角量aopをマップ演算する(S24)。ここで、マップ演算は、たとえば、入力変数の値がマップデータの入力変数の値のいずれかに一致する場合、対応するマップデータの出力変数の値を演算結果とするのに対し、一致しない場合、マップデータに含まれる複数の出力変数の値の補間によって得られる値を演算結果とする処理とすればよい。

20

## 【0035】

そしてCPU72は、スロットルバルブ14に操作信号MS1を出力してスロットル開口度TAを操作するとともに、点火装置26に操作信号MS3を出力して点火時期を操作する(S26)。ここで、本実施形態では、スロットル開口度TAをスロットル開口度指令値TA\*にフィードバック制御することを例示することから、スロットル開口度指令値TA\*が同一の値であっても、操作信号MS1が互いに異なる信号となりうるものである。また、たとえば周知のノッキングコントロール(KCS)等がなされる場合、点火時期は、基準点火時期を遅角量aopにて遅角させた値がKCSにてフィードバック補正された値とされる。ここで、基準点火時期は、CPU72により、クランク軸28の回転速度NEおよび充填効率 に応じて可変設定される。なお、回転速度NEは、クランク角センサ84の出力信号Scrに基づきCPU72によって算出される。また、充填効率は、回転速度NEおよび吸入空気量Gaに基づきCPU72によって算出される。

30

## 【0036】

なお、CPU72はS26の処理が完了する場合、図2に示す一連の処理を一旦終了する。

図3に、上記マップデータDMを生成するシステムを示す。

## 【0037】

図3に示すように、本実施形態では、内燃機関10のクランク軸28にトルクコンバータ40および変速装置50を介してダイナモメータ100を機械的に連結する。そして内燃機関10を稼働させた際の様々な状態変数がセンサ群102によって検出され、検出結果が、マップデータDMを生成するコンピュータである生成装置110に入力される。なお、センサ群102には、図1に示した車両VC1が搭載するセンサ等が含まれる。

40

## 【0038】

生成装置110は、CPU112、ROM114、電氣的に書き換え可能な不揮発性メモリ(記憶装置116)、および周辺回路118を備えており、それらがローカルネットワーク119によって通信可能とされたものである。ここで、記憶装置116には、アクセル操作量PAと、スロットル開口度指令値TA\*および遅角量aopとの関係を規定するデータである関係規定データDRが記憶されている。ここで、関係規定データDRは、ノーマルモード用規定データDR1、スポーツモード用規定データDR2、およびコンフォートモード用規定データDR3を備える。また、ROM114には、強化学習によって

50



、関係規定データDRを学習する学習プログラム114aが記憶されている。

【0039】

図4に、生成装置110が実行する処理の手順を示す。図4に示す処理は、ROM114に記憶された学習プログラム114aを、CPU112が実行することにより実現される。

【0040】

図4に示す一連の処理において、CPU112は、まず、モード変数MVの値を設定する(S30)。次にCPU112は、内燃機関10を稼働させた状態において、状態sとして、アクセル操作量PAの時系列データを取得する(S32)。ここでの時系列データは、S10の処理におけるものと同様のデータである。ただし、図4に示す処理において、アクセルペダル86は存在しない。そのため、アクセル操作量PAを、生成装置110が車両VC1の状態を模擬することによって疑似的に生成されたものとし、疑似的に生成されたアクセル操作量PAを、車両の状態の検出値とみなす。

10

【0041】

次にCPU112は、関係規定データDRのうちのS30の処理によって設定されたモード変数MVの値に対応するデータが定める方策に従い、S32の処理によって取得した状態sに応じたスロットル開口度指令値TA\*および遅角量aopからなる行動aを設定する(S34)。

【0042】

本実施形態において、関係規定データDRは、行動価値関数Qおよび方策を定めるデータである。本実施形態において、行動価値関数Qは、状態sおよび行動aの8次元の独立変数に応じた期待収益の値を示すテーブル型式の関数である。また、方策は、状態sが与えられたときに、独立変数が与えられた状態sとなる行動価値関数Qのうち最大となる行動a(グリーディ行動)を優先的に選択しつつも、所定の確率で、それ以外の行動aを選択する規則を定める。

20

【0043】

詳しくは、本実施形態にかかる行動価値関数Qの独立変数がとりうる値の数は、状態sおよび行動aのとりうる値の全組み合わせの一部が、人の知見等によって削減されたものである。すなわち、たとえばアクセル操作量PAの時系列データのうち隣接する2つのサンプリング値の1つがアクセル操作量PAの最小値となりもう1つが最大値となるようなことは、人によるアクセルペダル86の操作からは生じえないとして、行動価値関数Qが定義されていない。本実施形態では、人の知見等に基づく次元削減によって、行動価値関数Qを定義する状態sのとりうる値を、10の4乗個以下、より望ましくは10の3乗個以下に制限する。

30

【0044】

次にCPU112は、設定されたスロットル開口度指令値TA\*および遅角量aopに基づき、S26の処理と同様にして操作信号MS1, MS3を出力する(S36)。次にCPU112は、内燃機関10のトルクTrq、内燃機関10に対するトルク指令値Trq\*、および加速度Gxを取得する(S38)。ここで、CPU112は、トルクTrqを、ダイナモメータ100が生成する負荷トルクと変速装置50の変速比とに基づき算出する。また、トルク指令値Trq\*は、アクセル操作量PAに応じて設定される。また、CPU112は、加速度Gxを、ダイナモメータ100の負荷トルク等に基づき、仮に内燃機関10等が車両に搭載されていた場合に車両に生じると想定される値として算出する。すなわち、本実施形態においては、加速度Gxについても仮想的なものであるが、この加速度Gxについても、車両の状態の検出値であるとみなす。

40

【0045】

次にCPU72は、過渡フラグFが「1」であるか否かを判定する(S40)。過渡フラグFは、「1」である場合に過渡運転時であることを示し、「0」である場合に過渡運転時ではないことを示す。CPU112は、過渡フラグFが「0」であると判定する場合(S40:NO)、アクセル操作量PAの単位時間当たりの変化量PAの絶対値が所定

50

量  $Path$ 以上であるか否かを判定する (S42)。ここで、変化量  $PA$ は、たとえば、S42の処理の実行タイミングにおける最新のアクセル操作量  $PA$ と、同タイミングに対して単位時間だけ前におけるアクセル操作量  $PA$ との差とすればよい。

【0046】

CPU112は、所定量  $Path$ 以上であると判定する場合 (S42: YES)、過渡フラグ  $F$ に「1」を代入する (S44) 一方、所定量  $Path$ 未満であると判定する場合 (S42: NO)、S32の処理に戻る。

【0047】

これに対し、CPU112は、過渡フラグ  $F$ が「1」であると判定する場合 (S40: YES)、S44の処理がなされてから所定期間が経過したか否かを判定する (S46)。ここで、所定期間は、アクセル操作量  $PA$ の単位時間当たりの変化量  $PA$ の絶対値が所定量  $Path$ よりも小さい規定量以下となる状態が所定時間継続するまでの期間とする。CPU112は、所定期間が経過したと判定する場合 (S46: YES)、過渡フラグ  $F$ に「0」を代入する (S48) 一方、所定期間が経過していないと判定する場合 (S46: NO)、S32の処理に戻る。

【0048】

CPU112は、S44、S48の処理が完了する場合、1つのエピソードが終了したとして、強化学習によって行動価値関数  $Q$ を更新する (S50)。

図5に、S50の処理の詳細を示す。

【0049】

図5に示す一連の処理において、CPU112は、直近に終了されたエピソード中のトルク指令値  $Trq^*$ 、トルク  $Trq$ および加速度  $Gx$ の3つのサンプリング値の組からなる時系列データと、状態  $s$ および行動  $a$ の時系列データと、を取得する (S60)。ここで、直近のエピソードは、S48の処理に続いてS60の処理がなされる場合には、過渡フラグ  $F$ が継続して「1」となっていた期間であり、S44の処理に続いてS60の処理がなされる場合には、過渡フラグ  $F$ が継続して「0」となっていた期間である。

【0050】

図5には、カッコの中の数字が異なるものが、異なるサンプリングタイミングにおける変数の値であることを示す。たとえば、トルク指令値  $Trq^*(1)$ とトルク指令値  $Trq^*(2)$ とは、サンプリングタイミングが互いに異なるものである。また、直近のエピソードに属する行動  $a$ の時系列データを、行動集合  $A_j$ とし、同エピソードに属する状態  $s$ の時系列データを、状態集合  $S_j$ と定義する。

【0051】

次にCPU112は、直近のエピソードに属する任意のトルク  $Trq$ とトルク指令値  $Trq^*$ との差の絶対値が規定量  $Trq$ 以下である旨の条件 (ア)と、加速度  $Gx$ が下限値  $GxL$ 以上であって上限値  $GxH$ 以下である旨の条件 (イ)との論理積が真であるか否かを判定する (S62)。

【0052】

ここで、CPU112は、規定量  $Trq$ を、エピソードの開始時におけるアクセル操作量  $PA$ の単位時間当たりの変化量  $PA$ およびモード変数  $MV$ の値によって可変設定する。すなわち、CPU112は、エピソードの開始時におけるアクセル操作量  $PA$ の単位時間当たりの変化量  $PA$ に基づき過渡時に関するエピソードであると判定する場合、定常時と判定する場合と比較して、規定量  $Trq$ を大きい値に設定する。また、CPU112は、スポーツモードの場合に規定量  $Trq$ を最も小さくし、コンフォードモードの場合に規定量  $Trq$ を最も大きくする。

【0053】

また、CPU112は、下限値  $GxL$ を、エピソードの開始時におけるアクセル操作量  $PA$ の単位時間当たりの変化量  $PA$ およびモード変数  $MV$ の値によって可変設定する。すなわち、CPU112は、過渡時に関するエピソードであって且つ変化量  $PA$ が正である場合には、定常時に関するエピソードの場合と比較して、下限値  $GxL$ を大きい値に

10

20

30

40

50

設定する。また、CPU112は、過渡時に関するエピソードであって且つ変化量  $PA$  が負である場合には、定常時に関するエピソードの場合と比較して、下限値  $G \times L$  を小さい値に設定する。また、CPU112は、変化量  $PA$  が正であるなら、スポーツモードの場合に下限値  $G \times L$  を最も大きくし、コンフォードモードの場合に下限値  $G \times L$  を最も小さくする。また、CPU112は、変化量  $PA$  が負であるなら、スポーツモードの場合に下限値  $G \times L$  を最も小さくし、コンフォードモードの場合に下限値  $G \times L$  を最も大きくする。

#### 【0054】

また、CPU112は、上限値  $G \times H$  を、エピソードの開始時におけるアクセル操作量  $PA$  の単位時間当たりの変化量  $PA$  およびモード変数  $MV$  の値によって可変設定する。すなわち、CPU112は、過渡時に関するエピソードであって且つ変化量  $PA$  が正である場合には、定常時に関するエピソードの場合と比較して、上限値  $G \times H$  を大きい値に設定する。また、CPU112は、過渡時に関するエピソードであって且つ変化量  $PA$  が負である場合には、定常時に関するエピソードの場合と比較して、上限値  $G \times H$  を小さい値に設定する。また、CPU112は、変化量  $PA$  が正であるなら、スポーツモードの場合に上限値  $G \times H$  を最も大きくし、コンフォードモードの場合に上限値  $G \times H$  を最も小さくする。また、CPU112は、変化量  $PA$  が負であるなら、スポーツモードの場合に上限値  $G \times H$  を最も小さくし、コンフォードモードの場合に上限値  $G \times H$  を最も大きくする。

#### 【0055】

CPU112は、論理積が真であると判定する場合 (S62: YES)、報酬  $r$  に「10」を代入する一方 (S64)、偽であると判定する場合 (S62: NO)、報酬  $r$  に「-10」を代入する (S66)。CPU112は、S64、S66の処理が完了する場合、図3に示した記憶装置116に記憶されている関係規定データ  $DR$  のうち、S30の処理において設定されているモード変数  $MV$  の値に対応したデータを更新する。本実施形態では、ソフト方策オン型モンテカルロ法を用いる。

#### 【0056】

すなわち、CPU112は、上記S60の処理によって読み出した各状態と対応する行動との組によって定まる収益  $R(S_j, A_j)$  に、それぞれ、報酬  $r$  を加算する (S68)。ここで、「 $R(S_j, A_j)$ 」は、状態集合  $S_j$  の要素の1つを状態とし行動集合  $A_j$  の要素の1つを行動とする収益  $R$  を総括した記載である。次に、上記S60の処理によって読み出した各状態と対応する行動との組によって定まる収益  $R(S_j, A_j)$  のそれぞれについて、平均化して対応する行動価値関数  $Q(S_j, A_j)$  に代入する (S70)。ここで、平均化は、S68の処理がなされた回数によって、S68の処理によって算出された収益  $R$  を除算する処理とすればよい。なお、収益  $R$  の初期値はゼロとすればよい。

#### 【0057】

次にCPU112は、上記S60の処理によって読み出した状態について、それぞれ、対応する行動価値関数  $Q(S_j, A)$  のうち、最大値となるときのスロットル開口度指令値  $TA^*$  および遅角量  $aop$  の組である行動を、行動  $A_j^*$  に代入する (S72)。ここで、「 $A$ 」は、とりうる任意の行動を示す。なお、行動  $A_j^*$  は、上記S60の処理によって読み出した状態の種類に応じて各別の値となるものであるが、ここでは、表記を簡素化して、同一の記号にて記載している。

#### 【0058】

次に、CPU72は、上記S60の処理によって読み出した状態のそれぞれについて、対応する方策  $(A_j | S_j)$  を更新する (S74)。すなわち、行動の総数を、「 $|A|$ 」とすると、S72によって選択された行動  $A_j^*$  の選択確率を、「 $1 - \frac{1}{|A|}$ 」とする。また、行動  $A_j^*$  以外の「 $|A| - 1$ 」個の行動の選択確率を、それぞれ「 $\frac{1}{|A|}$ 」とする。S74の処理は、S70の処理によって更新された行動価値関数  $Q$  に基づく処理であることから、これにより、状態  $s$  と行動  $a$  との関係を規定する関係規定データ  $DR$  が、収益  $R$  を増加させるように更新されることとなる。

10

20

30

40

50

## 【 0 0 5 9 】

なお、CPU 112は、S74の処理が完了する場合、図5に示す一連の処理を一旦終了する。

図4に戻り、CPU 112は、S50の処理が完了すると、行動価値関数Qが収束したか否かを判定する(S52)。ここでは、S50の処理による行動価値関数Qの更新量が所定値以下となる連続回数が所定回数に達する場合に収束したと判定すればよい。CPU 112は、収束していないと判定する場合(S52:NO)、S32の処理に戻る。これに対し、CPU 112は、収束したと判定する場合(S52:YES)、モード変数MVの全ての値について、S52の処理によって肯定判定されたか否かを判定する(S54)。そして、CPU 112は、未だ行動価値関数Qが収束していないモード変数MVの値があると判定する場合(S54:NO)、そのモード変数MVの値に対応する行動価値関数Qの学習をすべく、S30の処理に戻る。

10

## 【 0 0 6 0 】

なお、CPU 112は、S54の処理において肯定判定する場合には、図4に示す一連の処理を一旦終了する。

図6に、生成装置110が実行する処理のうち、特に図4の処理によって学習された行動価値関数Qに基づき、マップデータDMを生成する処理の手順を示す。図6に示す処理は、ROM 114に記憶された学習プログラム114aを、CPU 112が実行することにより実現される。

## 【 0 0 6 1 】

20

図6に示す一連の処理において、CPU 112は、まず、モード変数MVの値を1つ設定する(S80)。次に、CPU 112は、マップデータDMの入力変数の値となる複数の状態sの中の1つを選択する(S82)。次に、CPU 112は、状態sに対応して且つ、S80の処理によって設定されたモードに関する行動価値関数Q(s,A)のうち、行動価値関数Qの値を最大とする行動aを選択する(S84)。すなわち、ここでは、グリーディ方策によって行動aを選択する。次に、CPU 112は、状態sと行動aとの組を、S80の処理によって設定されたモード変数MVの値に対応付けて記憶装置116に記憶させる(S86)。

## 【 0 0 6 2 】

次にCPU 112は、マップデータDMの入力変数の値とするもの全てがS80の処理によって選択されたか否かを判定する(S88)。そして、CPU 112は、選択されていないものがあると判定する場合(S88:NO)、S82の処理に戻る。これに対し、CPU 112は、全てが選択されたと判定する場合(S88:YES)、モード変数MVがとりうる値の全てについて、S88の処理において肯定判定されたか否かを判定する(S90)。そしてCPU 112は、未だS88の処理において肯定判定されていないモード変数MVの値があると判定する場合(S90:NO)、S80の処理に戻って、そのモード変数MVの値を設定する。これに対し、CPU 112は、全てのモード変数MVについてS88の処理において肯定判定されたと判定する場合(S90:YES)、S86の処理によって記憶されたデータに基づき、マップデータDMを生成する(S92)。ここでは、マップデータDMの入力変数の値が状態sであるものに対応する出力変数の値を、

30

40

## 【 0 0 6 3 】

なお、CPU 112は、S92の処理が完了する場合、図6に示す一連の処理を一旦終了する。

ここで、本実施形態の作用および効果について説明する。

## 【 0 0 6 4 】

図3に示すシステムにおいて、CPU 112は、強化学習によって、行動価値関数Qを学習する。そして、行動価値関数Qの値が収束する場合、走行制御モードを満たすうえで適切な行動が学習されたとする。そして、CPU 112は、マップデータDMの入力変数となる状態のそれぞれについて、行動価値関数Qを最大化する行動を選択し、状態と行動

50

との組を記憶装置 116 に記憶する。次に、CPU 112 は、記憶装置 116 に記憶された状態と行動との組に基づき、マップデータ DM を生成する。これにより、アクセル操作量 PA に応じた適切なスロットル開口度指令値 TA\* および遅角量 aop を、熟練者による工数を過度に大きくすることなく設定することができる。

#### 【0065】

特に、本実施形態では、モード変数 MV の値毎に、報酬 r の与え方を変えた。具体的には、CPU 112 は、たとえば、スポーツモードにおいてトルク Trq とトルク指令値 Trq\* との差の絶対値がノーマルモード用の規定量 Trq 以下であったとしても、スポーツモード用の規定量 Trq よりも大きい場合には、負の報酬を与える。これにより、スポーツモードにおいては、ノーマルモードと比較して、アクセル操作量 PA に応じたトルク指令値 Trq\* に実際のトルク Trq をより早期に追従させることができる行動 a が大きい報酬をうることができることとなり、車両の応答性能を高めることが可能な行動 a に対応する行動価値関数 Q の値が大きい値に更新される。

10

#### 【0066】

したがって、モードに応じて適切な関係規定データ DR1 ~ DR3 を学習できる。

以上説明した本実施形態によれば、さらに以下に記載する作用効果が得られる。

(1) 制御装置 70 が備える記憶装置 76 に、行動価値関数 Q 等ではなく、マップデータ DM を記憶した。これにより、CPU 72 は、マップデータ DM を用いたマップ演算に基づき、スロットル開口度指令値 TA\* や遅角量 aop を設定することから、行動価値関数 Q のうち最大値となるものを選択する処理を実行する場合と比較して、演算負荷を軽減できる。

20

#### 【0067】

(2) 行動価値関数 Q の独立変数にアクセル操作量 PA の時系列データを含めた。これにより、アクセル操作量 PA に関して単一のサンプリング値のみを独立変数とする場合と比較して、アクセル操作量 PA の様々な変化に対して行動 a の値をきめ細かく調整できる。

#### 【0068】

(3) 行動価値関数 Q の独立変数に、スロットル開口度指令値 TA\* 自体を含めた。これにより、たとえば、スロットル開口度指令値 TA\* の挙動をモデル化したモデル式のパラメータ等をスロットル開口度に関する独立変数とする場合と比較して、強化学習による探索の自由度を高めることが容易である。

30

#### 【0069】

< 第 2 の実施形態 >

以下、第 2 の実施形態について、第 1 の実施形態との相違点を中心に図面を参照しつつ説明する。

#### 【0070】

図 7 に、本実施形態にかかる車両 VC1 の駆動系および制御装置を示す。なお、図 7 において、図 1 に示した部材に対応する部材については、便宜上、同一の符号を付している。

#### 【0071】

図 7 に示すように、本実施形態では、ROM 74 に、制御プログラム 74a に加えて、学習プログラム 74b が記憶されている。また、記憶装置 76 に、マップデータ DM が記憶されておらず、代わりに、関係規定データ DR が記憶され、また、トルク出力写像データ DT が記憶されている。ここで、関係規定データ DR は、図 4 の処理によって学習された学習済みのデータである。また、トルク出力写像データ DT によって規定されるトルク出力写像は、回転速度 NE、充填効率、および点火時期を入力とし、トルク Trq を出力するニューラルネットワーク等の学習済みモデルに関するデータである。なお、上記トルク出力写像データ DT は、たとえば図 4 の処理を実行する際、S38 の処理によって取得されるトルク Trq を教師データとして学習されたものとすればよい。

40

#### 【0072】

50

図 8 に、本実施形態にかかる制御装置 70 が実行する処理の手順を示す。図 8 に示す処理は、ROM 74 に記憶された制御プログラム 74 a および学習プログラム 74 b を、CPU 72 がたとえば所定周期で繰り返し実行することにより実現される。なお、図 8 において、図 4 に示した処理に対応する処理については、便宜上同一のステップ番号を付与する。

#### 【0073】

図 8 に示す一連の処理において、CPU 72 は、まずモード選択スイッチ 92 の操作に応じたモード変数 MV の値を取得する (S30a)。そして、CPU 72 は、図 4 の S32 ~ S50 の処理と同様の処理を実行する。なお、CPU 72 は、S42, S46 の処理において否定判定する場合や、S50 の処理を完了する場合には、図 8 に示す一連の処理を一旦終了する。ちなみに、S30a, S32 ~ 48 の処理は、CPU 72 が制御プログラム 74 a を実行することにより実現され、S50 の処理は、CPU 72 が学習プログラム 74 b を実行することにより実現される。

10

#### 【0074】

このように、本実施形態によれば、制御装置 70 に関係規定データ DR および学習プログラム 74 b を実装することにより、第 1 の実施形態の場合と比較して、学習頻度を向上させることができる。

#### 【0075】

< 第 3 の実施形態 >

以下、第 3 の実施形態について、第 2 の実施形態との相違点を中心に図面を参照しつつ説明する。

20

#### 【0076】

本実施形態では、関係規定データ DR の更新を、車両 VC1 の外で実行する。

図 9 に、本実施形態において、強化学習を実行する制御システムの構成を示す。なお、図 9 において、図 1 に示した部材に対応する部材については、便宜上、同一の符号を付している。

#### 【0077】

図 9 に示す車両 VC1 内の制御装置 70 における ROM 74 は、制御プログラム 74 a を記憶しているものの、学習プログラム 74 b を記憶していない。また、制御装置 70 は、通信機 77 を備えている。通信機 77 は車両 VC1 の外部のネットワーク 120 を介してデータ解析センター 130 と通信するための機器である。

30

#### 【0078】

データ解析センター 130 は、複数の車両 VC1, VC2, ... から送信されるデータを解析する。データ解析センター 130 は、CPU 132、ROM 134、および電氣的に書き換え可能な揮発性メモリ (記憶装置 136)、周辺回路 138 および通信機 137 を備えており、それらがローカルネットワーク 139 によって通信可能とされるものである。ROM 134 には、学習プログラム 74 b が記憶されており、記憶装置 136 には、関係規定データ DR が記憶されている。

#### 【0079】

図 10 に、本実施形態にかかる強化学習の処理手順を示す。図 10 (a) に示す処理は、図 9 に示す ROM 74 に記憶されている制御プログラム 74 a を CPU 72 が実行することにより実現される。また、図 10 (b) に示す処理は、ROM 134 に記憶されている学習プログラム 74 b を CPU 132 が実行することにより実現される。なお、図 10 において図 8 に示した処理に対応する処理については、便宜上同一のステップ番号を付している。以下では、強化学習の時系列に沿って、図 10 に示す処理を説明する。

40

#### 【0080】

図 10 (a) に示す一連の処理において、CPU 72 は、S32 ~ S48 の処理を実行した後、通信機 77 を操作することによって、関係規定データ DR の更新処理に必要なデータを送信する (S100)。ここで、送信対象とされるデータは、S44, S48 の処理の実行直前に終了したエピソードにおけるモード変数 MV の値や、トルク指令値 Trq

50

\*、トルク  $T r q$  および加速度  $G x$  の時系列データ、状態集合  $S j$  および行動集合  $A j$  を含む。

【 0 0 8 1 】

これに対し、図 10 ( b ) に示すように、CPU 1 3 2 は、送信されたデータを受信し ( S 1 1 0 )、受信したデータに基づき関係規定データ DR を更新する ( S 5 0 )。そして CPU 1 3 2 は、関係規定データ DR の更新回数が所定回数以上であるか否かを判定し ( S 1 1 2 )、所定回数以上であると判定する場合 ( S 1 1 2 : Y E S )、通信機 1 3 7 を操作して、S 1 1 0 の処理によって受信したデータを送信した車両 VC 1 に関係規定データ DR を送信する ( S 1 1 4 )。なお、CPU 1 3 2 は、S 1 1 4 の処理を完了する場合や、S 1 1 2 の処理において否定判定する場合には、図 10 ( b ) に示す一連の処理を一旦終了する。

10

【 0 0 8 2 】

これに対し、図 10 ( a ) に示すように、CPU 7 2 は、更新データがあるか否かを判定し ( S 1 0 2 )、あると判定する場合 ( S 1 0 2 : Y E S )、更新された関係規定データ DR を受信する ( S 1 0 4 )。そして CPU は、S 3 4 の処理において利用する関係規定データ DR を、受信した関係規定データ DR に書き換える ( S 1 0 6 )。なお、CPU 7 2 は、S 1 0 6 の処理を完了する場合や、S 1 0 2 の処理において否定判定する場合には、図 10 ( a ) に示す一連の処理を一旦終了する。

【 0 0 8 3 】

このように、本実施形態によれば、関係規定データ DR の更新処理を車両 VC 1 の外部で行うことから、制御装置 7 0 の演算負荷を軽減できる。

20

< 対応関係 >

上記実施形態における事項と、上記「課題を解決するための手段」の欄に記載した事項との対応関係は、次の通りである。以下では、「課題を解決するための手段」の欄に記載した解決手段の番号毎に、対応関係を示している。

【 0 0 8 4 】

[ 1 , 4 ] 実行装置と記憶装置とは、図 1 および図 7 においては、それぞれ、CPU 7 2 および ROM 7 4 と記憶装置 7 6 とに対応し、図 3 においては、それぞれ CPU 1 1 2 および ROM 1 1 4 と記憶装置 1 1 6 とに対応し、図 9 においては、CPU 7 2 , 1 3 2 および ROM 7 4 , 1 3 4 と記憶装置 7 6 , 1 3 6 とに対応する。操作処理は、S 3 6 の処理に対応し、取得処理は、S 3 2 , S 3 8 の処理に対応する。報酬算出処理は、S 6 2 ~ S 6 6 の処理に対応する。更新処理は、S 6 8 ~ S 7 4 の処理に対応する。走行制御モードの相違は、モード変数 MV によって定まる。[ 2 ] S 6 2 の処理に対応する。[ 3 ] 開口度に関する変数は、スロットル開口度指令値  $T A *$  に対応する。[ 5 ] 時系列データは、アクセル操作量  $P A ( 1 ) , P A ( 2 ) , \dots P A ( 6 )$  に対応する。[ 6 ] 制御用写像データは、マップデータ DM に対応する。実行装置は、CPU 1 1 2 および ROM 1 1 4 に対応する。[ 7 ] 実行装置と記憶装置とは、それぞれ、図 7 における CPU 7 2 および ROM 7 4 と記憶装置 7 6 とに対応する。取得処理は、S 3 0 a , S 3 2 , S 3 8 の処理に対応する。[ 8 ~ 1 0 ] 第 1 実行装置は、CPU 7 2 および ROM 7 4 に対応し、第 2 実行装置は、CPU 1 3 2 および ROM 1 3 4 に対応する。

30

40

【 0 0 8 5 】

< その他の実施形態 >

なお、本実施形態は、以下のように変更して実施することができる。本実施形態および以下の変更例は、技術的に矛盾しない範囲で互いに組み合わせて実施することができる。

【 0 0 8 6 】

・「走行制御モードについて」

上記実施形態では、アクセル操作に対する車両の応答性が高い順に、スポーツモード、ノーマルモード、およびコンフォートモードの 3 つの走行制御モードのうちの 1 つを選択可能な車両を例示したが、これに限らない。たとえば、応答性について、3 種類の性能に代えて、2 種類の性能から選択可能としたり、4 種類以上の性能から選択可能としたりし

50

てもよい。

【0087】

また、車両の走行制御モードとしては、応答性の観点からの性能に限らない。たとえば燃料消費率やエネルギー消費率の観点からの性能としてもよい。

・「行動変数について」

上記実施形態では、行動変数としてのスロットルバルブの開口度に関する変数として、スロットル開口度指令値 $T A *$ を例示したが、これに限らない。たとえば、アクセル操作量 $P A$ に対するスロットル開口度指令値 $T A *$ の応答性を、無駄時間および2次遅れフィルタにて表現し、無駄時間と、2次遅れフィルタを規定する2つの変数との合計3つの変数を、スロットルバルブの開口度に関する変数としてもよい。ただし、その場合、状態変数は、アクセル操作量 $P A$ の時系列データに代えて、アクセル操作量 $P A$ の単位時間当たりの変化量とすることが望ましい。

10

【0088】

上記実施形態では、行動変数としての点火時期に関する変数として、遅角量 $a o p$ を例示したが、これに限らない。たとえば、 $K C S$ による補正対象とされる点火時期自体であってもよい。

【0089】

上記実施形態では、行動変数として、スロットルバルブの開口度に関する変数および点火時期に関する変数を例示したが、これに限らない。たとえば、スロットルバルブの開口度に関する変数および点火時期に関する変数に加えて、燃料噴射量を用いてもよい。また、それら3つに関しては、行動変数としてスロットルバルブの開口度に関する変数および燃料噴射量のみを採用したり、点火時期に関する変数および燃料噴射量のみを採用したりしてもよい。さらに、それら3つに関しては、行動変数としてそれらのうちの1つのみを採用してもよい。

20

【0090】

また、「内燃機関について」の欄に記載したように、圧縮着火式の内燃機関の場合、スロットルバルブの開口度に関する変数に代えて噴射量に関する変数を用い、点火時期に関する変数に代えて噴射時期に関する変数を用いればよい。なお、噴射時期に関する変数に加えて、1燃焼サイクルにおける噴射回数に関する変数や、1燃焼サイクルにおける1つの気筒のための時系列的に隣接する2つの燃料噴射のうち一方の終了タイミングと他方の開始タイミングとの間の時間間隔に関する変数を加えることが望ましい。

30

【0091】

また、たとえば変速装置 $50$ が有段変速装置の場合、クラッチの係合状態を油圧によって調整するためのソレノイドバルブの電流値等を行動変数としてもよい。

また、たとえば、下記「車両について」の欄に記載したように車両としてハイブリッド車や、電気自動車、燃料電池車を採用する場合、回転電機のトルクや出力を行動変数としてもよい。またたとえば、内燃機関のクランク軸の回転動力によって回転するコンプレッサを備えた車載空調装置を備える場合、コンプレッサの負荷トルクを行動変数に含めてもよい。また、電動式の車載空調装置を備える場合、空調装置の消費電力を行動変数に含めてもよい。

40

【0092】

・「状態について」

上記実施形態では、アクセル操作量 $P A$ の時系列データを、等間隔でサンプリングされた6個の値からなるデータとしたが、これに限らない。互いに異なるサンプリングタイミングにおける2個以上のサンプリング値からなるデータであればよく、この際、3個以上のサンプリング値からなるデータや、サンプリング間隔が等間隔であるデータであることがより望ましい。

【0093】

アクセル操作量に関する状態変数としては、アクセル操作量 $P A$ の時系列データに限らず、たとえば「行動変数について」の欄に記載したように、アクセル操作量 $P A$ の単位時

50



間当たりの変化量等であってもよい。

【 0 0 9 4 】

また、たとえば「行動変数について」の欄に記載したように、ソレノイドバルブの電流値を行動変数とする場合、状態に、変速装置の入力軸 5 2 の回転速度や出力軸 5 4 の回転速度、ソレノイドバルブによって調整される油圧を含めればよい。またたとえば「行動変数について」の欄に記載したように、回転電機のトルクや出力を行動変数とする場合、状態に、バッテリーの充電率や温度を含めればよい。またたとえば「行動変数について」の欄に記載したように、コンプレッサの負荷トルクや空調装置の消費電力を行動に含める場合、状態に、車室内の温度を含めればよい。

【 0 0 9 5 】

- ・「テーブル形式のデータの次元削減について」

テーブル形式のデータの次元削減手法としては、上記実施形態において例示したものに限らない。たとえばアクセル操作量  $PA$  が最大値となることはまれであることから、アクセル操作量  $PA$  が規定量以上となる状態については行動価値関数  $Q$  を定義せず、アクセル操作量  $PA$  が規定量以上となる場合のスロットル開口度指令値  $TA^*$  等は、別途適合してもよい。またたとえば、行動のとりうる値からスロットル開口度指令値  $TA^*$  が規定値以上となるものを除くなどして、次元削減をしてもよい。

【 0 0 9 6 】

図 4 の  $S 5 2$  の処理において肯定判定されるまでは、 $S 3 4$  の処理において、行動価値関数  $Q$  の独立変数の値を少数に制限し、 $S 5 2$  の処理において肯定判定される場合に、行動価値関数  $Q$  の値が大きくなる行動  $a$  の付近の値を行動価値関数  $Q$  の独立変数がとりうる値に加えて  $S 3 2 \sim S 5 2$  の処理を繰り返してもよい。

【 0 0 9 7 】

もっとも、次元削減をすることは必須ではない。たとえば、第 3 の実施形態において複数の車両からのデータに基づき強化学習を行って且つ  $CPU 7 2$  の演算能力や記憶装置 7 6 の記憶容量が十分であるのであれば、車両の出荷前には次元削減をした一部のみについて行動価値関数を学習しておくものの、出荷後には、全ての行動を探索によって実行可能としてもよい。これにより、出荷後には出荷前と比較して十分な学習用のデータを確保できることに鑑み、探索としてとりうる行動の数を増やして、より適切な行動を見出すことが可能となる。

【 0 0 9 8 】

- ・「関係規定データについて」

上記実施形態では、行動価値関数  $Q$  を、テーブル形式の関数としたが、これに限らない。たとえば、関数近似器を用いてもよい。

【 0 0 9 9 】

たとえば、行動価値関数  $Q$  を用いる代わりに、方策  $\pi$  を、状態  $s$  および行動  $a$  を独立変数とし、行動  $a$  をとる確率を従属変数とする関数近似器にて表現し、関数近似器を定めるパラメータを、報酬  $r$  に応じて更新してもよい。

【 0 1 0 0 】

- ・「操作処理について」

たとえば「関係規定データについて」の欄に記載したように、行動価値関数を関数近似器とする場合、上記実施形態におけるテーブル形式の関数の独立変数となる行動についての離散的な値の組の全てについて、状態  $s$  とともに行動価値関数  $Q$  に入力することによって、行動価値関数  $Q$  を最大化する行動  $a$  を選択すればよい。

【 0 1 0 1 】

また、たとえば「関係規定データについて」の欄に記載したように、方策  $\pi$  を、状態  $s$  および行動  $a$  を独立変数とし、行動  $a$  をとる確率を従属変数とする関数近似器とする場合、方策  $\pi$  によって示される確率に基づき行動  $a$  を選択すればよい。

【 0 1 0 2 】

- ・「更新写像について」

10

20

30

40

50

S 6 8 ~ S 7 4 の処理においては、ソフト方策オン型モンテカルロ法によるものを例示したが、これに限らない。たとえば、方策オフ型モンテカルロ法によるものであってもよい。もっとも、モンテカルロ法にも限らず、たとえば、方策オフ型TD法を用いたり、またたとえばSARS A法のように方策オン型TD法を用いたり、またたとえば、方策オン型の学習として適格度トレース法を用いたりしてもよい。

【0103】

また、たとえば「関係規定データについて」の欄に記載したように、方策を関数近似器を用いて表現し、これを報酬  $r$  に基づき直接更新する場合には、方策勾配法等を用いて更新写像を構成すればよい。

【0104】

また、行動価値関数  $Q$  と方策  $\pi$  のうちのいずれか一方のみを、報酬  $r$  による直接の更新対象とするものに限らない。たとえば、アクター・クリティック法のように、行動価値関数  $Q$  および方策  $\pi$  をそれぞれ更新してもよい。また、アクター・クリティック法においては、これに限らず、たとえば行動価値関数  $Q$  に代えて価値関数  $V$  を更新対象としてもよい。

【0105】

なお、方策  $\pi$  を定める「 $\pi$ 」については、固定値に限らず、学習の進行度合いに応じてあらかじめ定められた規則に応じて変更してもよい。

・「報酬算出処理について」

図5の処理では、条件(ア)および条件(イ)の論理積が真であるか否かに応じて報酬を与えたが、これに限らない。たとえば、条件(ア)を満たすか否かに応じて報酬を与える処理と、条件(イ)を満たすか否かに応じて報酬を与える処理とを実行してもよい。また、たとえば、条件(ア)を満たすか否かに応じて報酬を与える処理と、条件(イ)を満たすか否かに応じて報酬を与える処理との2つの処理に関しては、それらのうちのいずれか1つの処理のみを実行してもよい。

【0106】

また、たとえば条件(ア)を満たす場合に一律同じ報酬を与える代わりに、トルク  $T r q$  とトルク指令値  $T r q *$  との差の絶対値が小さい場合に大きい場合よりもより大きい報酬を与える処理としてもよい。またたとえば、条件(ア)を満たさない場合に一律同じ報酬を与える代わりに、トルク  $T r q$  とトルク指令値  $T r q *$  との差の絶対値が大きい場合に小さい場合よりもより小さい報酬を与える処理としてもよい。そして、それらの場合、与える報酬の大きさを、モード変数  $M V$  の値に応じて変更してもよい。

【0107】

また、たとえば条件(イ)を満たす場合に一律同じ報酬を与える代わりに、加速度  $G x$  の大きさに応じて報酬の大きさを可変とする処理としてもよい。またたとえば、条件(イ)を満たさない場合に一律同じ報酬を与える代わりに、加速度  $G x$  の大きさに応じて報酬の大きさを可変とする処理としてもよい。そして、それらの場合、与える報酬の大きさを、モード変数  $M V$  の値に応じて変更してもよい。

【0108】

また、たとえば「行動変数について」の欄に記載したように、変速装置50のソレノイドバルブの電流値を行動変数とする場合、たとえば報酬算出処理に以下の(a)~(c)の3つの処理のうちの少なくとも1つの処理を含めればよい。

【0109】

(a) 変速装置による変速比の切り替えに要する時間が所定時間以内である場合に所定時間を超える場合よりも大きい報酬を与える処理である。

(b) 変速装置の入力軸52の回転速度の変化速度の絶対値が入力側所定値以下である場合に入力側所定値を超える場合よりも大きい報酬を与える処理である。

【0110】

(c) 変速装置の出力軸54の回転速度の変化速度の絶対値が出力側所定値以下である場合に入力側所定値を超える場合よりも大きい報酬を与える処理である。

10

20

30

40

50

こうした場合において、たとえばスポーツモードのように応答性を重視した走行制御モードが選択される場合には、上記所定時間をより短く設定する一方、入力側所定値や出力側所定値についてはより大きい値に設定すればよい。

【0111】

また、たとえば「走行制御モードについて」の欄に記載したように、燃料消費率やエネルギー消費率の観点から走行制御モードが定義されている場合においては、燃料消費率やエネルギー消費率が低い場合に高い場合よりも大きい報酬を与える処理とすればよい。また、この際、「行動変数について」の欄に記載したように、車載空調装置を備える場合には、その負荷トルクや消費電力を行動変数に含めてもよい。

【0112】

また、たとえば「行動変数について」の欄に記載したように、回転電機のトルクや出力を行動変数とする場合、バッテリーの充電率が所定範囲内にある場合でない場合よりも大きい報酬を与える処理や、バッテリーの温度が所定範囲内にある場合でない場合よりも大きい報酬を与える処理を含めてもよい。また、たとえば「行動変数について」の欄に記載したように、コンプレッサの負荷トルクや空調装置の消費電力を行動変数に含める場合、車室内の温度が所定範囲内にある場合でない場合よりも大きい報酬を与える処理を加えてもよい。

【0113】

- ・「車両用制御データの生成方法について」

図4のS34の処理では、行動価値関数Qに基づき行動を決定したが、これに限らず、とりうるすべての行動を等確率で選択してもよい。

【0114】

- ・「制御用写像データについて」

車両の状態と期待収益を最大化する行動変数の値とを1対1に対応付けることによって車両の状態を入力とし期待収益を最大化する行動変数の値を出力する制御用写像データとしては、マップデータに限らない。たとえば、関数近似器であってもよい。これは、たとえば、上記「更新写像について」の欄に記載したように、方策勾配法等を用いる場合において、方策を行動変数の値をとりうる確率を示すガウス分布にて表現し、その平均値を関数近似器にて表現しておき、平均値を表現する関数近似器のパラメータを更新することとし、学習後の平均値を制御用写像データとすることによって実現できる。

【0115】

- ・「取得処理について」

上記実施形態では、ユーザによる走行制御モードの選択結果を、モード選択スイッチ92の出力信号に基づくモード変数MVを取得することによって取得したがこれに限らない。たとえば、モード選択スイッチ92に代えて、音声指示を感知する装置を備え、その感知結果をモード変数MVとして取得してもよい。

【0116】

- ・「車両用制御システムについて」

図10に示した例では、方策に基づく行動を決定する処理(S34の処理)を、車両側で実行したが、これに限らない。たとえば、車両VC1からS32の処理によって取得したデータを送信することとし、データ解析センター130にて送信されてデータを用いて行動aを決定し、決定した行動を車両VC1に送信してもよい。

【0117】

車両用制御システムとしては、制御装置70およびデータ解析センター130によって構成されるものに限らない。たとえば、データ解析センター130に代えて、ユーザの携帯端末を用いてもよい。また、制御装置70およびデータ解析センター130と携帯端末とによって車両用制御システムを構成してもよい。これは、たとえばS34の処理を携帯端末によって実行することにより実現できる。

【0118】

- ・「実行装置について」

実行装置としては、CPU72(112, 132)とROM74(114, 134)とを備えて、ソフトウェア処理を実行するものに限らない。たとえば、上記実施形態においてソフトウェア処理されたものの少なくとも一部を、ハードウェア処理するたとえばASIC等の専用のハードウェア回路を備えてもよい。すなわち、実行装置は、以下の(a)~(c)のいずれかの構成であればよい。(a)上記処理の全てを、プログラムに従って実行する処理装置と、プログラムを記憶するROM等のプログラム格納装置とを備える。(b)上記処理の一部をプログラムに従って実行する処理装置およびプログラム格納装置と、残りの処理を実行する専用のハードウェア回路とを備える。(c)上記処理の全てを実行する専用のハードウェア回路を備える。ここで、処理装置およびプログラム格納装置を備えたソフトウェア実行装置や、専用のハードウェア回路は複数であってもよい。

10

## 【0119】

・「記憶装置について」

上記実施形態では、関係規定データDRが記憶される記憶装置と、学習プログラム74b, 114aや制御プログラム74aが記憶される記憶装置(ROM74, 114, 134)とを別の記憶装置としたが、これに限らない。

## 【0120】

・「内燃機関について」

内燃機関としては、燃料噴射弁として吸気通路12に燃料を噴射するポート噴射弁を備えるものに限らず、燃烧室24に燃料を直接噴射する筒内噴射弁を備えるものであってもよく、またたとえば、ポート噴射弁および筒内噴射弁の双方を備えるものであってもよい。

20

## 【0121】

内燃機関としては、火花点火式内燃機関に限らず、たとえば燃料として軽油などを用いる圧縮着火式内燃機関等であってもよい。

・「車両について」

車両としては、推力生成装置が内燃機関のみである車両に限らず、たとえば内燃機関と回転電機とを備えるいわゆるハイブリッド車両であってもよい。またたとえば、推力生成装置として、内燃機関を備えることなく、回転電機を備えるいわゆる電気自動車や燃料電池車であってもよい。

30

## 【符号の説明】

## 【0122】

10...内燃機関、12...吸気通路、14...スロットルバルブ、16...燃料噴射弁、18...吸気バルブ、20...シリンダ、22...ピストン、24...燃烧室、26...点火装置、28...クランク軸、30...排気バルブ、32...排気通路、34...触媒、40...トルクコンバータ、42...ロックアップクラッチ、50...変速装置、52...入力軸、54...出力軸、60...駆動輪、70...制御装置、72...CPU、74...ROM、74a...制御プログラム、74b...学習プログラム、76...記憶装置、77...通信機、78...周辺回路、79...ローカルネットワーク、80...エアフローメータ、82...スロットルセンサ、84...クランク角センサ、86...アクセルペダル、88...アクセルセンサ、90...加速度センサ、92手mンモード選択スイッチ、100...ダイナモメータ、102...センサ群、110...生成装置、112...CPU、114...ROM、114a...学習プログラム、116...記憶装置、118...周辺回路、119...ローカルネットワーク、120...ネットワーク、130...データ解析センター、132...CPU、134...ROM、136...記憶装置、137...通信機、138...周辺回路、139...ローカルネットワーク。

40

## 【要約】

【課題】車両の状態と行動変数との関係の設定に際して熟練者に要求される工数を削減できるようにした車両用制御データの生成方法を提供する。

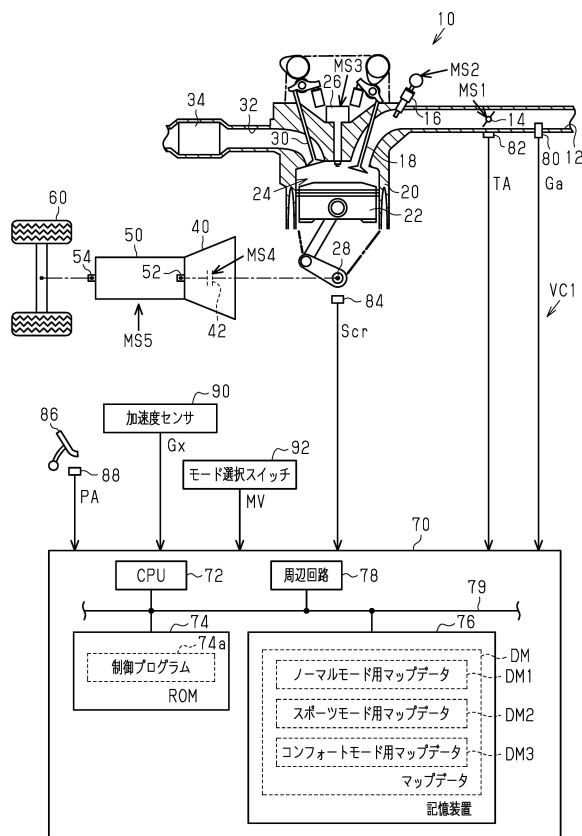
【解決手段】CPUは、1つのエピソードが終了するまで(S44, S48)、アクセル操作量PAの時系列データに基づき、スロットル開口度指令値TA\*および点火時期の遅角量aopを設定し、それらに応じてスロットルバルブや点火装置を操作し(S36)、

50

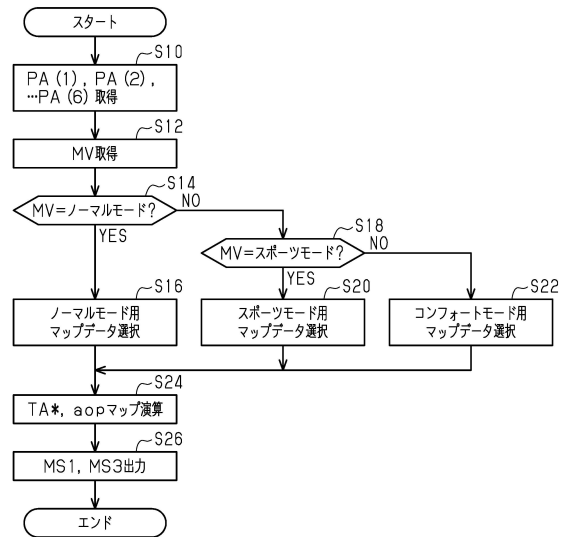
その際のトルク  $T r q$ 、トルク指令値  $T r q *$  および加速度  $G x$  を取得する ( S 3 8 )。CPU は、エピソードが終了する場合、トルク  $T r q$  や加速度  $G x$  が基準を満たすか否かに応じて報酬を与えることによって行動価値関数  $Q$  を更新する ( S 5 0 )。CPU は、走行モードに応じて、報酬の与え方を変える。

【選択図】図4

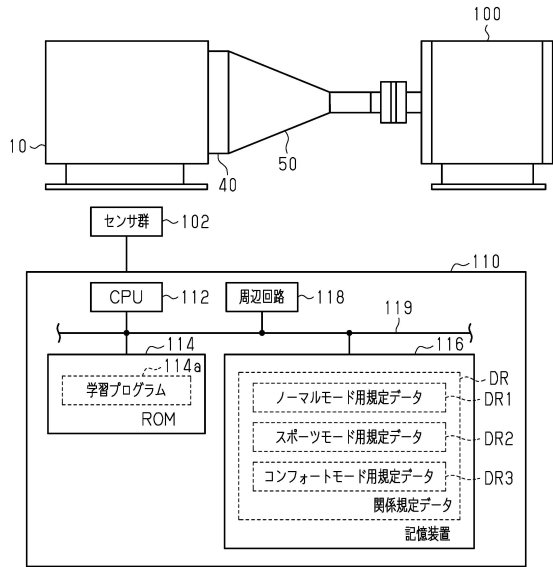
【図1】



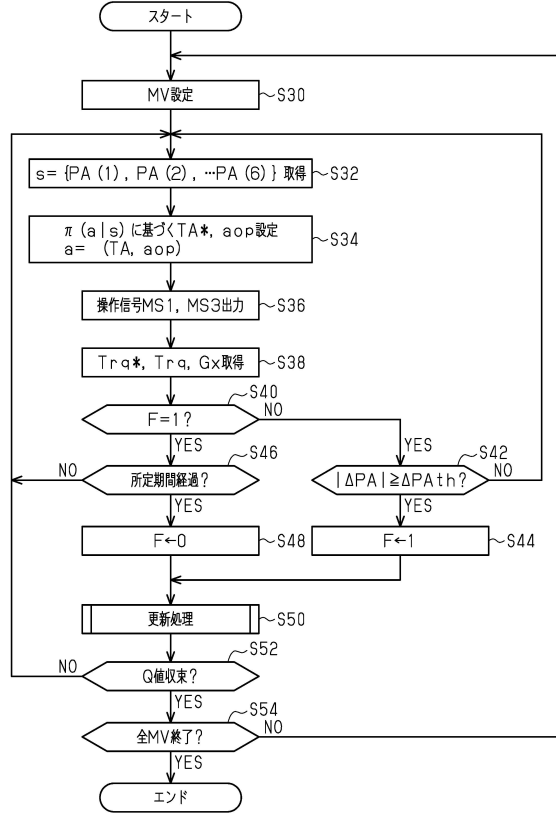
【図2】



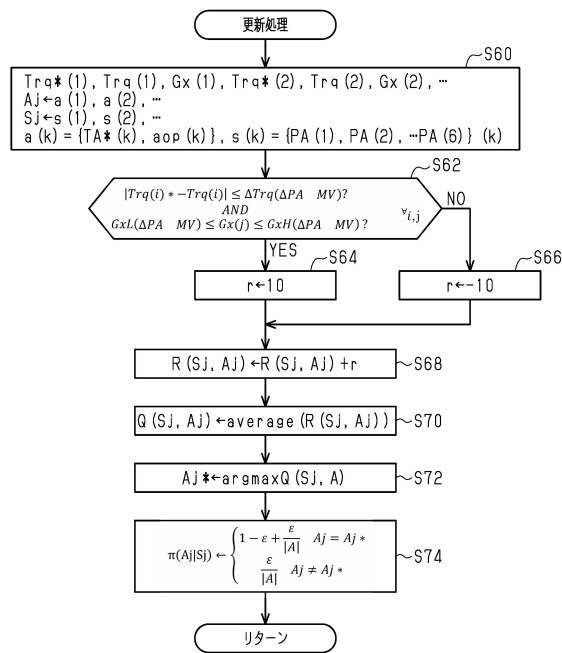
【図3】



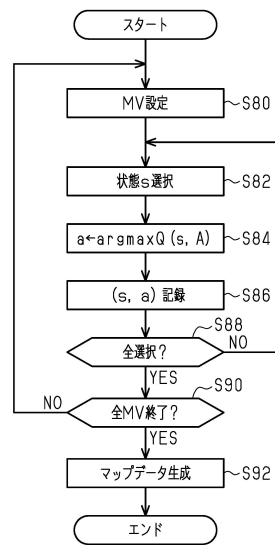
【図4】



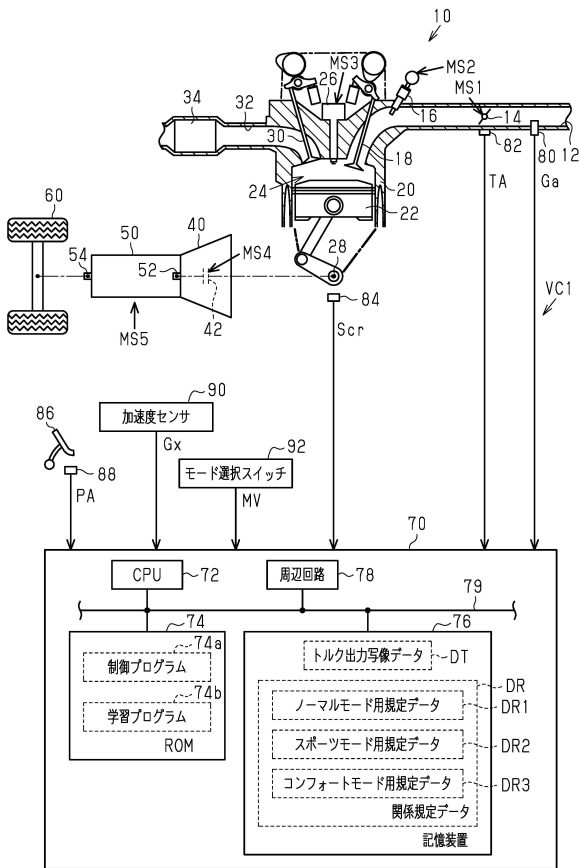
【図5】



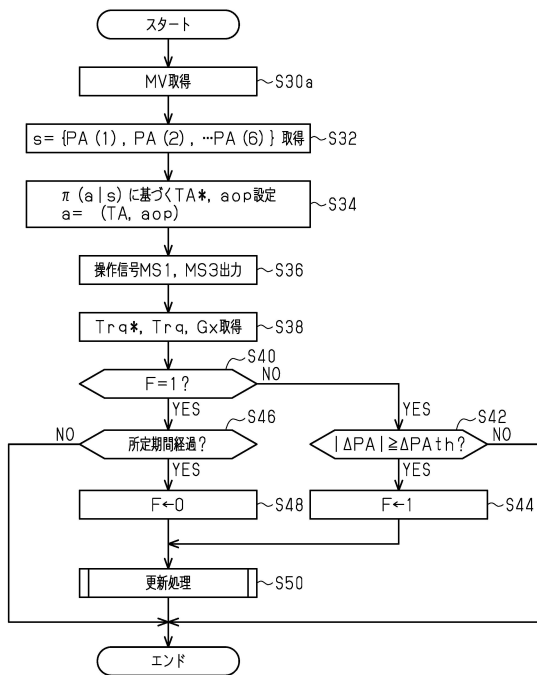
【図6】



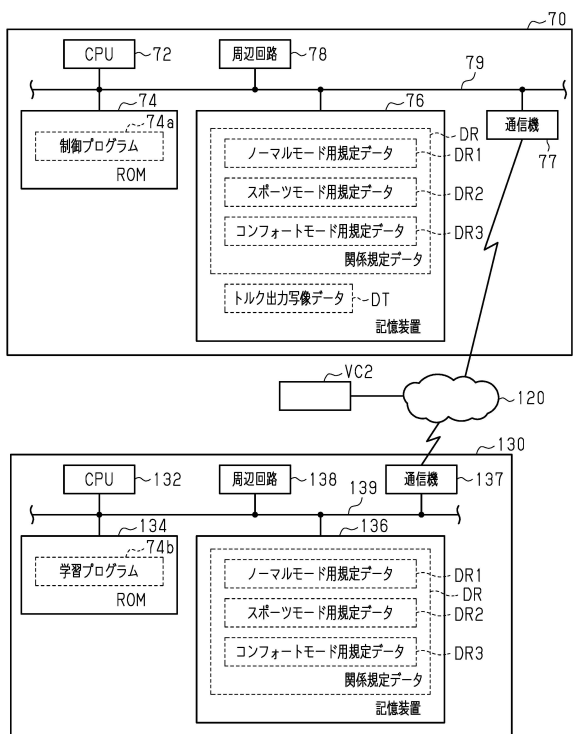
【図7】



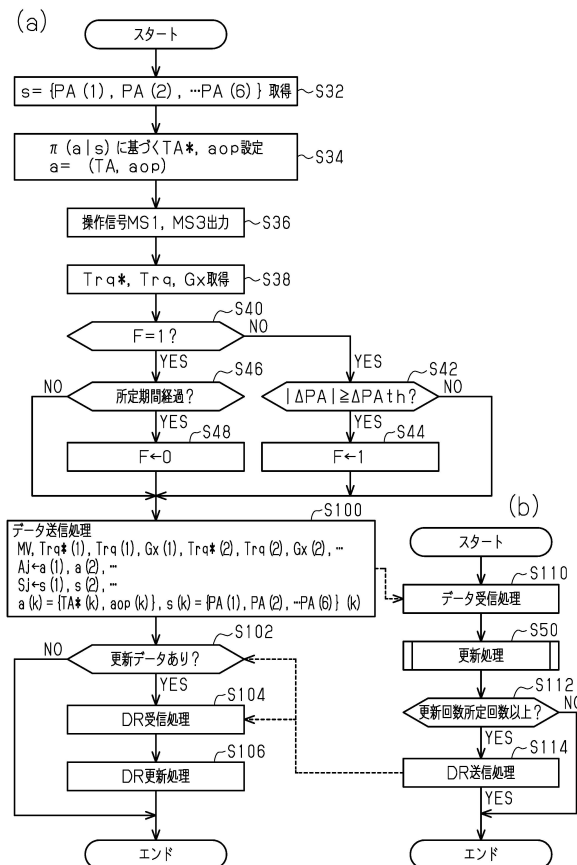
【図8】



【図9】



【図10】



## フロントページの続き

(51)Int.Cl. F I  
G 1 6 Y 20/20 (2020.01) G 1 6 Y 20/20  
G 1 6 Y 40/20 (2020.01) G 1 6 Y 40/20

(72)発明者 大城 裕太  
愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内  
(72)発明者 杉江 和紀  
愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内  
(72)発明者 岡 尚哉  
愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内

審査官 佐々木 淳

(56)参考文献 特表2019-519851(JP,A)  
特開2000-250604(JP,A)  
特開平10-254505(JP,A)  
特開2019-144748(JP,A)  
特開平04-293626(JP,A)  
特表平11-515070(JP,A)  
特開2017-91272(JP,A)

(58)調査した分野(Int.Cl., DB名)  
F 0 2 D 9 / 0 0 - 4 5 / 0 0  
B 6 0 W 1 0 / 0 0 - 5 0 / 1 6  
G 0 5 B 1 1 / 3 6  
G 0 6 N 2 0 / 0 0 - 9 9 / 0 0