

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2021-67261
(P2021-67261A)

(43) 公開日 令和3年4月30日(2021.4.30)

(51) Int.Cl.	F I	テーマコード (参考)
FO2D 45/00 (2006.01)	FO2D 45/00 364	3G301
FO2D 41/04 (2006.01)	FO2D 45/00 362	3G384
B6OR 16/02 (2006.01)	FO2D 41/04	
	B6OR 16/02 660T	

審査請求 未請求 請求項の数 5 O L (全 17 頁)

(21) 出願番号 特願2020-67631 (P2020-67631)
 (22) 出願日 令和2年4月3日(2020.4.3)
 (62) 分割の表示 特願2019-191094 (P2019-191094)の分割
 原出願日 令和1年10月18日(2019.10.18)

(71) 出願人 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

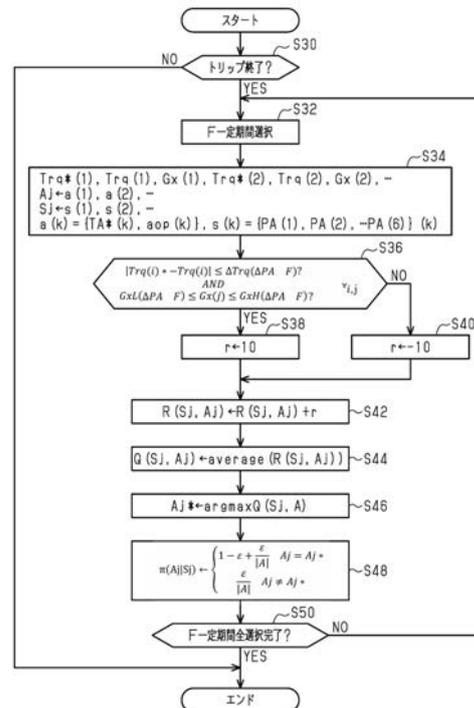
(54) 【発明の名称】 車両用制御装置、車両用制御システム、および車両用学習装置

(57) 【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装置を提供する。

【解決手段】CPUは、トリップの終了時であると判定する場合(S30: YES)、同トリップ中における、状態としてのアクセル操作量PAの時系列データ、行動としてのスロットル開度指令値TA*および点火時期の遅角量aop、報酬を算出するためのトルクTrq、トルク指令値Trq*および加速度Gxを読み出す(S34)。CPUは、トルクTrqや加速度Gxが基準を満たすか否かに応じて報酬を与える(S36~S40)。CPUは、与えた報酬に基づき、行動価値関数Qおよび方策を更新する(S42~S48)。

【選択図】図3



【特許請求の範囲】**【請求項 1】**

実行装置および記憶装置を備え、

前記記憶装置には、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されており、

前記実行装置は、

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と前記関係規定データとによって定まる前記行動変数の値に基づき前記電子機器を操作する操作処理と、

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前記行動変数の値、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、

を実行し、

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、前記操作処理、当該第 1 実行装置の演算負荷が所定以下であるときに前記更新処理に必要なデータを送信する車両側送信処理、および前記更新処理によって更新された前記関係規定データを受信する車両側受信処理を実行し、

前記第 2 実行装置は、少なくとも前記車両側送信処理によって送信されたデータを受信する外部側受信処理、前記更新処理、および前記更新処理によって更新された前記関係規定データを前記車両に送信する外部側送信処理を実行する車両用制御システム。

【請求項 2】

前記車両は、内燃機関を備え、

前記操作処理は、前記内燃機関の操作部を操作する処理を含み、

前記第 1 実行装置は、前記内燃機関のクランク軸の回転速度が所定速度以下であるときに前記演算負荷が所定以下であるとして、前記車両側送信処理を実行する請求項 1 記載の車両用制御システム。

【請求項 3】

前記第 1 実行装置は、前記車両が停車しているときに前記演算負荷が所定以下であるとして、前記車両側送信処理を実行する請求項 1 または 2 記載の車両用制御システム。

【請求項 4】

請求項 1 ~ 3 のいずれか 1 項に記載の車両用制御システムにおける前記第 1 実行装置を備える車両用制御装置。

【請求項 5】

請求項 1 ~ 3 のいずれか 1 項に記載の車両用制御システムにおける前記第 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】****【0001】**

本発明は、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

10

20

30

40

50

載されている。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2016-6327号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

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】

【0005】

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. 実行装置および記憶装置を備え、前記記憶装置には、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されており、前記実行装置は、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と前記関係規定データとによって定まる前記行動変数の値に基づき前記電子機器を操作する操作処理と、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前記行動変数の値、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、を実行し、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、前記実行装置は、前記更新処理を、当該実行装置の演算負荷が所定以下であるときに実行する車両用制御装置である。

20

【0006】

上記構成では、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車両の走行において適切な関係に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

【0007】

ところで、更新処理を実行することにより、実行装置の演算負荷が大きくなる。そこで上記構成では、演算負荷が所定以下のときに更新処理を実行することにより、実行装置が実行すべき他のタスクに更新処理を実行することによる影響が及ぶことを抑制できる。

【0008】

2. 前記車両は、内燃機関を備え、前記操作処理は、前記内燃機関の操作部を操作する処理を含み、前記実行装置は、前記内燃機関のクランク軸の回転速度が所定速度以下であるときに前記演算負荷が所定以下であるとして、前記更新処理を実行する上記1記載の車両用制御装置。

40

【0009】

内燃機関の制御量を制御すべくその操作部を操作する処理は、圧縮上死点の出現間隔に応じた処理を含むことから、クランク軸の回転速度が高い場合に低い場合よりも、内燃機関の制御のための演算負荷が大きくなる。そこで上記構成では、回転速度が所定速度以下であるときに更新処理を実行することにより、内燃機関の制御量の制御に関する演算負荷と更新処理の演算負荷とによって実行装置の演算負荷が過大となることを抑制できる。

50

【 0 0 1 0 】

3. 前記実行装置は、前記車両が停車しているときに前記演算負荷が所定以下であるとして、前記更新処理を実行する上記1または2記載の車両用制御装置である。

車両の走行時には、停車時と比較して実行装置の演算負荷が大きくなる傾向がある。そこで上記構成では、車両が停車しているときに更新処理を実行することにより、車両の走行に伴って実行装置が実行する処理による演算負荷と更新処理による演算負荷とによって実行装置の演算負荷が過大となることを抑制できる。

【 0 0 1 1 】

4. 上記1～3のいずれか1つに記載の前記取得処理、前記操作処理、前記報酬算出処理、および前記更新処理を実行する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、前記操作処理、当該第1実行装置の演算負荷が前記所定以下であるときに前記更新処理に必要なデータを送信する車両側送信処理、および前記更新処理によって更新された前記関係規定データを受信する車両側受信処理を実行し、前記第2実行装置は、少なくとも前記車両側送信処理によって送信されたデータを受信する外部側受信処理、前記更新処理、および前記更新処理によって更新された前記関係規定データを前記車両に送信する外部側送信処理を実行する車両用制御システムである。

10

【 0 0 1 2 】

上記構成では、第2実行装置が更新処理を実行することにより、第1実行装置の演算負荷を軽減できる。しかも、第1実行装置が、演算負荷が所定以下であるときに車両側送信処理を実行することにより、車両側送信処理による演算負荷に起因して第1実行装置の演算負荷が過大となることを抑制できる。

20

【 0 0 1 3 】

5. 上記4記載の第1実行装置を備える車両用制御装置である。

6. 上記4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 】

【 0 0 1 4 】

【 図 1 】 第1の実施形態にかかる制御装置およびその駆動系を示す図。

【 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制御装置が実行する処理の手順を示す流れ図。

【 図 4 】 第2の実施形態にかかる車両用制御システムの構成を示す図。

【 図 5 】 (a) および (b) は、車両用制御システムが実行する処理の手順を示す流れ図。

30

【 発明を実施するための形態 】

【 0 0 1 5 】

< 第1の実施形態 >

以下、車両用制御装置にかかる第1の実施形態について、図面を参照しつつ説明する。

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。

【 0 0 1 6 】

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。

40

【 0 0 1 7 】

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

50

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。

【0018】

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。

10

【0019】

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 G_a 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 の開口度（スロットル開口度 T_A ）、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 G_x を参照する。

【0020】

制御装置 70 は、CPU 72、ROM 74、電气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。

20

【0021】

ROM 74 には、制御プログラム 74a および学習プログラム 74b 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 と、スロットル開口度 T_A の指令値（スロットル開口度指令値 T_A^* ）および点火装置 26 の遅角量 aop との関係の規定する関係規定データ DR 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。また、記憶装置 76 には、トルク出力写像データ DT が記憶されている。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、クランク軸 28 の回転速度 NE 、充填効率、および点火時期を入力とし、トルク Trq を出力する写像である。

30

【0022】

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、ROM 74 に記憶された制御プログラム 74a 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、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。

40

【0023】

図 2 に示す一連の処理において、CPU 72 は、まず、状態 s として、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。

【0024】

50

次にCPU72は、関係規定データDRが定める方策に従い、S10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(S12)。

【0025】

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形式の関数である。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。

10

【0026】

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うちの一部が、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。

20

【0027】

次にCPU72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、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(S14)。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、たとえば周知のロックアップコントロール(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

【0028】

次にCPU72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(S16)。ここで、CPU112は、トルクTrqを、回転速度NE、充填効率および点火時期をトルク出力写像に入力することによって算出する。また、CPU72は、トルク指令値Trq*を、アクセル操作量PAに応じて設定する。

【0029】

次にCPU72は、過渡フラグFが「1」であるか否かを判定する(S18)。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。CPU72は、過渡フラグFが「0」であると判定する場合(S18:NO)、アクセル操作量PAの単位時間当たりの変化量PAの絶対値が所定量PAth以上であるか否かを判定する(S20)。ここで、変化量PAは、たとえば、S20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。

40

【0030】

CPU72は、所定量PAth以上であると判定する場合(S20:YES)、過渡フラグFに「1」を代入する(S22)。

これに対し、CPU72は、過渡フラグFが「1」であると判定する場合(S18:Y

50

ES)、S22の処理の実行タイミングから所定期間が経過したか否かを判定する(S24)。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。CPU72は、所定期間が経過したと判定する場合(S24:YES)、過渡フラグFに「0」を代入する(S26)。

【0031】

CPU72は、S22, S26の処理が完了する場合や、S20, S24の処理において否定判定する場合には、S28の処理に移行する。CPU72は、S28の処理において、S10の処理によって取得した状態s、S12の処理によって設定した行動a、S16の処理によって取得したトルクTrq、トルク指令値Trq*および加速度Gx、および現在の過渡フラグFの値を、記憶装置76に記憶する。なお、CPU72は、S28の処理を完了する場合、図2に示す一連の処理を一旦終了する。

10

【0032】

図3に、本実施形態において制御装置70が実行する処理の手順を示す。図3に示す処理は、ROM74に記憶された学習プログラム74b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。

【0033】

図3に示す一連の処理において、CPU72は、まず、トリップの終了時であるか否かを判定する(S30)。ここで、トリップとは、車両の走行許可信号がオン状態である1回の期間のことである。本実施形態において、走行許可信号は、イグニッション信号に相当する。

20

【0034】

CPU72は、トリップの終了時であると判定する場合(S30:YES)、過渡フラグFが一定である1つの期間、すなわち1つのエピソードを選択する(S32)。各エピソードは、S26の処理がなされてからS22の処理がなされるまでの期間や、S22の処理がなされてからS26の処理がなされるまでの期間のことである。

【0035】

次に、CPU72は、S32の処理において選択した1つのエピソード内の、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、を以下の処理に利用するデータとして読み出す(S34)。図3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、選択したエピソードに属する行動aの時系列データを、行動集合Aj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Sjと定義する。

30

【0036】

次にCPU72は、選択した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(S36)。

40

【0037】

ここで、CPU72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の符号と過渡フラグFの値とによって可変設定する。すなわち、CPU72は、過渡フラグFが「1」であるエピソードの場合、過渡時に関するエピソードであるとして、過渡フラグFが「0」である場合と比較して、規定量Trqを大きい値に設定する。また、CPU72は、変化量PAの符号に応じて規定量Trqを互いに異なる値に設定する。

【0038】

また、CPU72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の変化量PAの符号と過渡フラグFの値とによって可変設定する。すなわち、CPU

50

72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 $G \times L$ を大きい値に設定する。また、CPU72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。

【0039】

また、CPU72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 の符号と過渡フラグ F の値とによって可変設定する。すなわち、CPU72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、CPU72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。

10

【0040】

CPU72は、論理積が真であると判定する場合 (S36: YES)、報酬 r に「10」を代入する一方 (S38)、偽であると判定する場合 (S36: NO)、報酬 r に「-10」を代入する (S40)。CPU72は、S38, S40の処理が完了する場合、図1に示した記憶装置76に記憶されている関係規定データ DR 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。

【0041】

すなわち、CPU72は、上記S34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 (S42)。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S32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 (S44)。ここで、平均化は、S42の処理がなされた回数に所定数を加えた数によって、S42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、行動価値関数 Q の初期値とすればよい。

20

【0042】

次にCPU72は、上記S34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 (S46)。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S34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。

30

【0043】

次に、CPU72は、上記S34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 (S48)。すなわち、行動の総数を、「 $|A|$ 」とすると、S46によって選択された行動 A_j^* の選択確率を、「 $1 - \dots / |A|$ 」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「 $\dots / |A|$ 」とする。S48の処理は、S44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。

40

【0044】

CPU72は、S48の処理が完了する場合、S28の処理によって記憶された全てのエピソードがS32の処理において選択され、S34~S48の処理が完了したか否かを判定する (S50)。CPU72は、未だ選択されていないエピソードがあると判定する場合 (S50: NO)、S32の処理に戻って、そのエピソードを選択する。これに対し、CPU72は、全てのエピソードが選択されたと判定する場合 (S50: YES) や、S30の処理において否定判定する場合には、図3に示す一連の処理を一旦終了する。

50

【0045】

ここで、本実施形態の作用および効果について説明する。

CPU72は、ユーザによるアクセルペダル86の操作に伴って、アクセル操作量PAの時系列データを取得し、方策に従って、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。ここでCPU72は、基本的には、関係規定データDRに規定されている行動価値関数Qに基づき期待収益を最大とする行動aを選択する。ただし、CPU72は、所定の確率で、期待収益を最大化する行動a以外の行動を選択することによって、期待収益を最大化する行動aの探索を行う。これにより、ユーザによる車両VC1の運転に伴って、関係規定データDRを強化学習によって更新できる。したがって、アクセル操作量PAに応じた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車両VC1の走行において適切な値に設定することができる。

10

【0046】

特に本実施形態では、更新処理を、トリップの終了時に行う。トリップの終了時には、トリップ中と比較して、内燃機関10の制御に関する演算負荷が小さいことから、CPU72の演算負荷が比較的小さい。そのため、CPU72によりS32~S50の処理を好適に実行できる。

【0047】

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) 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。

20

【0048】

(2) 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。

【0049】

<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。

30

【0050】

図4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4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。

【0051】

図4に示す車両VC1内のROM74には、制御プログラム74aに加えて、学習用サブプログラム74cが記憶されている。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100を介してデータ解析センター110と通信するための機器である。

40

【0052】

データ解析センター110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110は、CPU112、ROM114、および電氣的に書き換え可能な不揮発性メモリ(記憶装置116)、周辺回路118および通信機117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るものである。ROM114には、学習用メインプログラム114aが記憶されている。記憶装置116には、関係規定データDRが記憶されている。

【0053】

図5に、本実施形態にかかる強化学習の処理手順を示す。図5(a)に示す処理は、図

50

4 に示す ROM 7 4 に記憶された学習用サブプログラム 7 4 c を CPU 7 2 が実行することにより実現される。また、図 5 (b) に示す処理は、ROM 1 1 4 に記憶されている学習用メインプログラム 1 1 4 a を CPU 1 1 2 が実行することにより実現される。なお、図 5 において図 3 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 5 に示す処理を説明する。

【 0 0 5 4 】

図 5 (a) に示す一連の処理において、CPU 7 2 は、S 3 0 の処理において肯定判定する場合、通信機 7 7 を操作して、関係規定データ DR の更新に必要なデータを送信する (S 6 0) 。すなわち、CPU 7 2 は、このトリップにおいて S 2 8 の処理により記憶した、状態 s 、行動 a 、トルク Trq 、トルク指令値 Trq^* 、加速度 Gx 、過渡フラグ F の時系列データを送信する。

10

【 0 0 5 5 】

これに対し、図 5 (b) に示すように、CPU 1 1 2 は、S 6 0 の処理によって送信されたデータを受信し (S 7 0)、S 3 2 ~ S 5 0 の処理を実行する。そして、CPU 1 1 2 は、S 5 0 の処理において肯定判定する場合、通信機 1 1 7 を操作することによって、更新された関係規定データ DR を送信する (S 7 2) 。なお、CPU 1 1 2 は、S 7 2 の処理を完了する場合、図 5 (b) に示す一連の処理を一旦終了する。

【 0 0 5 6 】

これに対し、図 5 (a) に示すように、CPU 7 2 は、更新された関係規定データ DR を受信し (S 6 2)、このデータによって、S 1 2 の処理において用いる関係規定データ DR を上書きする (S 6 4) 。

20

【 0 0 5 7 】

なお、CPU 7 2 は、S 6 4 の処理が完了する場合や、S 3 0 の処理において否定判定する場合には、図 5 (a) に示す一連の処理を一旦終了する。

このように、本実施形態においては、関係規定データ DR の更新処理をデータ解析センター 1 1 0 において実行することにより、CPU 7 2 の演算負荷を軽減できる。

【 0 0 5 8 】

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。

(3) CPU 7 2 は、トリップの終了時に、関係規定データ DR の更新に必要なデータを送信した。これにより、トリップ中に送信する場合と比較してトリップ中に送信のために必要な演算負荷を軽減できる。

30

【 0 0 5 9 】

< 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 6 0 】

[1 ~ 3] 実行装置は、CPU 7 2 および ROM 7 4 に対応し、記憶装置は、記憶装置 7 6 に対応する。取得処理は、S 1 0、S 1 6 の処理に対応し、操作処理は、S 1 4 の処理に対応し、報酬算出処理は、S 3 6 ~ S 4 0 の処理に対応し、更新処理は、S 4 2 ~ S 4 8 の処理に対応する。更新写像は、学習プログラム 7 4 b のうち S 4 2 ~ S 4 8 の処理を実行する指令によって規定された写像に対応する。演算負荷が所定以下であるときは、トリップの終了時に対応する。[4 ~ 7] 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 1 2 および ROM 1 1 4 に対応する。車両側送信処理は、S 6 0 の処理に対応し、車両側受信処理は、S 6 2 の処理に対応する。外部側受信処理は、S 7 0 の処理に対応し、外部側送信処理は、S 7 2 の処理に対応する。

40

【 0 0 6 1 】

< 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。

50

【 0 0 6 2 】

・「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。

【 0 0 6 3 】

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。

【 0 0 6 4 】

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。

【 0 0 6 5 】

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。

【 0 0 6 6 】

また、たとえば変速装置 $5 0$ 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。

【 0 0 6 7 】

・「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。

【 0 0 6 8 】

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 $P A$ の単位時間当たりの変化量等であってもよい。

【 0 0 6 9 】

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

10

20

30

40

50

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 7 0 】

・「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。

10

【 0 0 7 1 】

たとえば、行動価値関数 Q を用いる代わりに、方策 π 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 0 7 2 】

・「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。

20

【 0 0 7 3 】

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 2 の実施形態において、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、車両の出荷前には行動価値関数の独立変数としてとりうる行動のうちの一部についてのみ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 7 4 】

・「更新写像について」

S 4 2 ~ S 4 8 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 S A R S 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。

30

【 0 0 7 5 】

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 π 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。

40

【 0 0 7 6 】

また、行動価値関数 Q と方策 π と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 π 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。

【 0 0 7 7 】

なお、方策 π を定める「 π 」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」

50

図3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。

【0078】

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。

10

【0079】

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。

【0080】

上記実施形態では、報酬 r を、ドライバビリティに関する基準を満たすか否かに応じて与えたが、ドライバビリティに関する基準としては、上述したものに限らず、たとえば騒音や振動強度が基準を満たすか否かに応じて設定してもよい。もっともこれに限らず、たとえば上記加速度が基準を満たすか否かと、トルク $T_r q$ の追従性が基準を満たすか否かと、騒音が基準を満たすか否かと、振動強度が基準を満たすか否かとの4つのうちの任意の1つ以上であってよい。

20

【0081】

報酬算出処理としては、報酬 r を、ドライバビリティに関する基準を満たすか否かに応じて与えるものにも限らない。たとえば、燃料消費率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であってもよい。またたとえば、排気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処理であってもよい。なお、ドライバビリティに関する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と、燃料消費率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と、排気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処理との3つの処理のうち2つまたは3つを含んでもよい。

30

【0082】

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つの処理のうち少なくとも1つの処理を含めればよい。

【0083】

(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。

40

(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。

【0084】

(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。

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

50

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。

【 0 0 8 5 】

・「車両用制御システムについて」

車両用制御システムとしては、制御装置 7 0 およびデータ解析センター 1 1 0 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 1 1 0 に代えて、ユーザが所持する携帯端末を用い、制御装置 7 0 および携帯端末によって車両用制御システムを構成してもよい。また、たとえば、制御装置 7 0、携帯端末、およびデータ解析センター 1 1 0 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 S 1 2 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。

10

【 0 0 8 6 】

・「実行装置について」

実行装置としては、CPU 7 2 (1 1 2) と ROM 7 4 (1 1 4) 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。

20

【 0 0 8 7 】

・「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 7 4 b や制御プログラム 7 4 a が記憶される記憶装置 (ROM 7 4) とを別の記憶装置としたが、これに限らない。

【 0 0 8 8 】

・「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 1 2 に燃料を噴射するポート噴射弁を備えるものに限らず、燃焼室 2 4 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。

30

【 0 0 8 9 】

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。

40

【符号の説明】

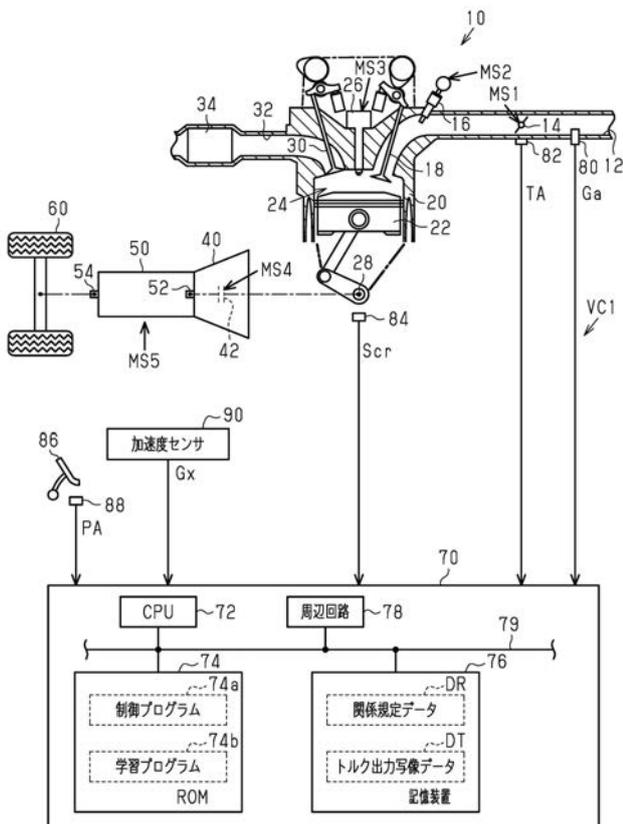
【 0 0 9 0 】

1 0 ... 内燃機関、 1 2 ... 吸気通路、 1 4 ... スロットルバルブ、 1 6 ... 燃料噴射弁、 1 8 ... 吸気バルブ、 2 0 ... シリンダ、 2 2 ... ピストン、 2 4 ... 燃焼室、 2 6 ... 点火装置、 2 8 ... クランク軸、 3 0 ... 排気バルブ、 3 2 ... 排気通路、 3 4 ... 触媒、 4 0 ... トルクコンバータ、 4 2 ... ロックアップクラッチ、 5 0 ... 変速装置、 5 2 ... 入力軸、 5 4 ... 出力軸、 6 0 ... 駆動輪、 7 0 ... 制御装置、 7 2 ... CPU、 7 4 ... ROM、 7 4 a ... 制御プログラム、 7 4 b ... 学習プログラム、 7 4 c ... 学習用サブプログラム、 7 6 ... 記憶装置、 7 7 ... 通信機、 7 8 ... 周辺回路、 7 9 ... ローカルネットワーク、 8 0 ... エアフローメータ、 8 2 ... スロ

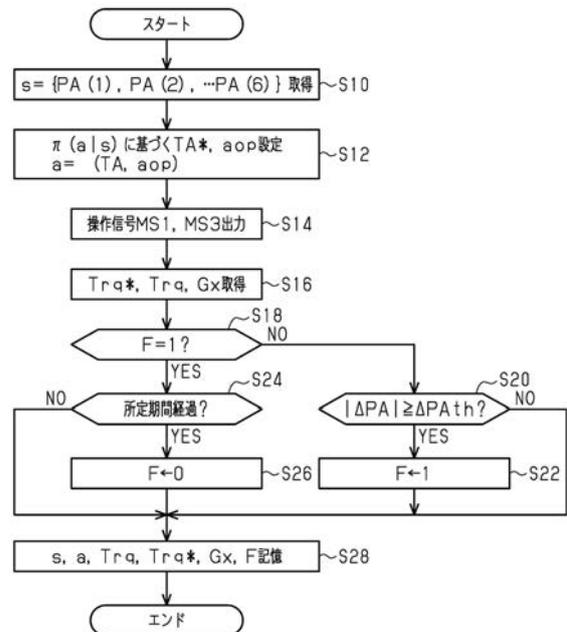
50

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、100...ネットワーク、110...データ解析センター、112...CPU、114...ROM、114a...学習用メインプログラム、116...記憶装置、117...通信機、118...周辺回路、119...ローカルネットワーク。

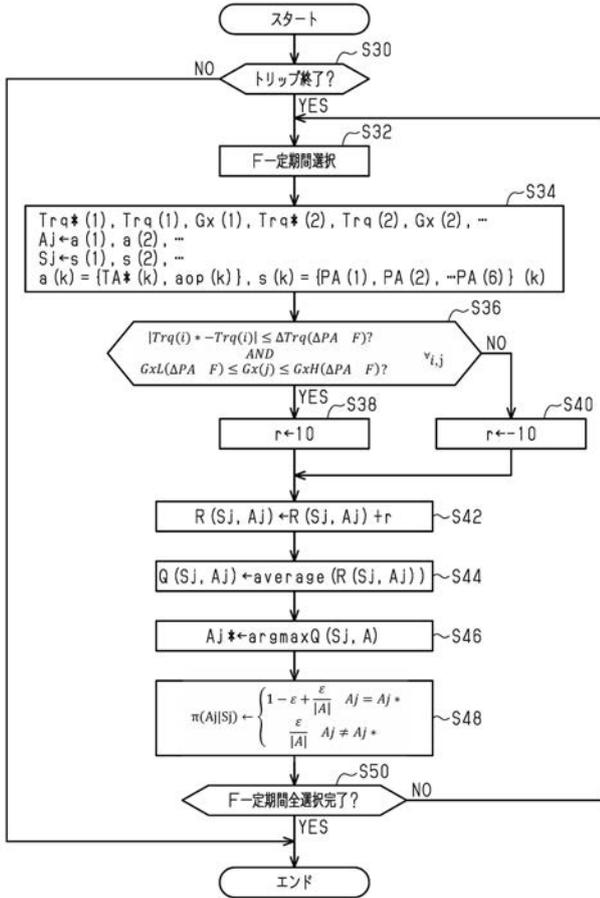
【 図 1 】



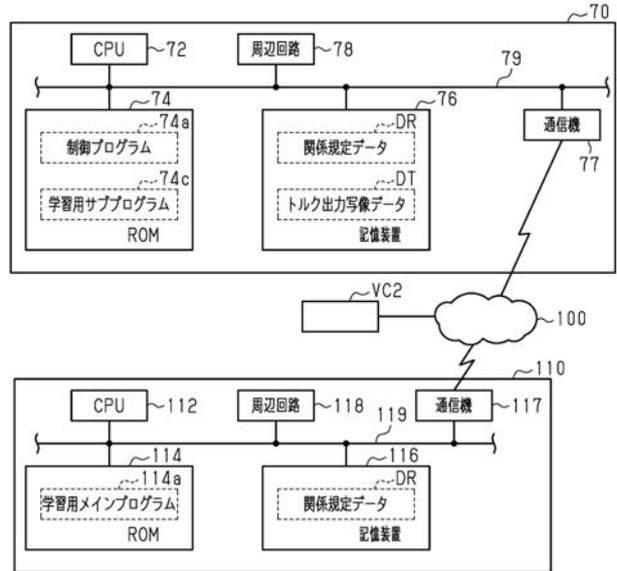
【 図 2 】



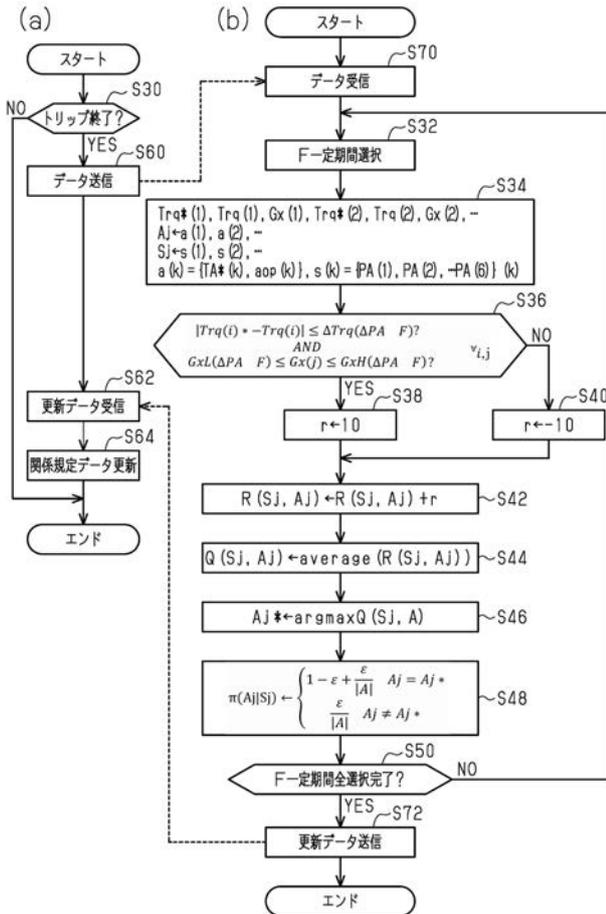
【 図 3 】



【 図 4 】



【 図 5 】



フロントページの続き

(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

Fターム(参考) 3G301 JA17 LA01 NA09 NB03 NB07 ND25 NE17 NE19 PE01Z PE09Z
PF02Z PF04Z PF13Z
3G384 AA28 BA02 BA05 BA13 BA18 BA24 DA18 EA11 EC01 EC05
EE17 FA07Z FA52Z FA56Z FA80Z FA81Z