

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6499720号
(P6499720)

(45) 発行日 平成31年4月10日 (2019. 4. 10)

(24) 登録日 平成31年3月22日 (2019. 3. 22)

(51) Int. Cl.

F I

G 0 5 B 19/404 (2006. 01)

G 0 5 B 19/404

G

G 0 5 B 13/02 (2006. 01)

G 0 5 B 13/02

L

請求項の数 7 (全 27 頁)

(21) 出願番号 特願2017-122449 (P2017-122449)
 (22) 出願日 平成29年6月22日 (2017. 6. 22)
 (65) 公開番号 特開2019-8472 (P2019-8472A)
 (43) 公開日 平成31年1月17日 (2019. 1. 17)
 審査請求日 平成30年8月22日 (2018. 8. 22)

早期審査対象出願

(73) 特許権者 390008235
 ファナック株式会社
 山梨県南都留郡忍野村忍草字古馬場 3 5 8
 O 番地
 (74) 代理人 100106002
 弁理士 正林 真之
 (74) 代理人 100165157
 弁理士 芝 哲央
 (74) 代理人 100160794
 弁理士 星野 寛明
 (72) 発明者 猪飼 聡史
 山梨県南都留郡忍野村忍草字古馬場 3 5 8
 O 番地 ファナック株式会社内

最終頁に続く

(54) 【発明の名称】 機械学習装置、サーボ制御装置、サーボ制御システム、及び機械学習方法

(57) 【特許請求の範囲】

【請求項 1】

位置指令又は位置偏差に対するバックラッシュ補正量を作成するバックラッシュ補正手段及び／又は、速度指令に対するバックラッシュ加速補正量を作成するバックラッシュ加速補正手段を備える、サーボ制御装置に対して、前記バックラッシュ補正手段におけるバックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正手段におけるバックラッシュ加速補正パラメータを対象とする強化学習を行う機械学習装置であって、

前記バックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正パラメータの調整情報を含む行動情報を前記バックラッシュ補正手段及び／又は前記バックラッシュ加速補正手段に対して出力する行動情報出力手段と、

前記行動情報に基づいて、所定の加工プログラムを前記サーボ制御装置に実行させた場合における、前記位置指令と位置フィードバックから得られる位置偏差と、前記バックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正パラメータと、を含む状態情報を前記サーボ制御装置から取得する状態情報取得手段と、

前記状態情報に含まれる前記位置偏差に基づいて、強化学習における報酬の値を出力する報酬出力手段と、

前記報酬の値と、前記状態情報と、前記行動情報とに基づいて行動価値関数を更新する価値関数更新手段と、

を備える機械学習装置。

【請求項 2】

10

20

前記バックラッシ補正手段におけるバックラッシ補正パラメータを対象として強化学習を行った後に、前記バックラッシ加速補正手段における前記バックラッシ加速補正パラメータを対象として強化学習を行う、請求項 1 に記載の機械学習装置。

【請求項 3】

前記報酬出力手段は、前記位置偏差の絶対値に基づいて前記報酬の値を出力する請求項 1 又は請求項 2 に記載の機械学習装置。

【請求項 4】

前記価値関数更新手段により更新された行動価値関数に基づいて、前記バックラッシ補正手段におけるバックラッシ補正パラメータ及び / 又は前記バックラッシ加速補正手段における前記バックラッシ加速補正パラメータを生成して出力する最適化行動情報出力手段を備えた請求項 1 から請求項 3 の何れか 1 項に記載の機械学習装置。

10

【請求項 5】

請求項 1 ～ 4 の何れか 1 項に記載の機械学習装置を含む前記サーボ制御装置。

【請求項 6】

請求項 1 ～ 4 の何れか 1 項に記載の機械学習装置と、前記サーボ制御装置と、を備えたサーボ制御システム。

【請求項 7】

位置指令又は位置偏差に対するバックラッシ補正量を作成するバックラッシ補正手段と、速度指令に対するバックラッシ加速補正量を作成するバックラッシ加速補正手段とを備える、サーボ制御装置に対して前記バックラッシ補正手段におけるバックラッシ補正パラメータ及び / 又は前記バックラッシ加速補正手段におけるバックラッシ加速補正パラメータを対象とする強化学習を行う機械学習装置の機械学習方法であって、

20

前記バックラッシ補正パラメータ及び / 又は前記バックラッシ加速補正パラメータの調整情報を含む行動情報を前記バックラッシ補正手段及び / 又は前記バックラッシ加速補正手段に対して出力する行動情報出力ステップと、

前記行動情報に基づいて、所定の加工プログラムを前記サーボ制御装置に実行させた場合における、前記位置指令と位置フィードバックから得られる位置偏差と、前記バックラッシ補正パラメータ及び / 又は前記バックラッシ加速補正パラメータと、を含む状態情報を前記サーボ制御装置から取得する状態情報取得ステップと、

前記状態情報に含まれる前記位置偏差に基づいて、強化学習における報酬の値を出力する報酬出力ステップと、

30

前記報酬の値と、前記状態情報と、前記行動情報とに基づいて行動価値関数を更新する価値関数更新ステップと、

を備える機械学習方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、バックラッシ補正制御及びバックラッシ加速補正制御を実行するサーボ制御装置に対して、バックラッシ補正制御におけるバックラッシ補正パラメータ及び / 又はバックラッシ加速補正制御におけるバックラッシ加速補正パラメータを対象とした強化学習を行う機械学習装置、当該機械学習装置を含むサーボ制御装置、及び機械学習装置とサーボ制御装置とを含むサーボ制御システム、並びに機械学習方法に関する。

40

【背景技術】

【0002】

従来、物体の位置、方位、姿勢等を制御量とする制御系、例えば工作機械や産業機械の制御系において、サーボモータが用いられている。サーボモータのモータ軸の回転を他の主軸に伝えるために用いられる機械要素であって互いに係合する機械要素においては、当該機械要素間の抵触面に意図的に隙間が設けられている。この隙間が存在することにより、例えばねじやギア等の機械要素は、ある程度の範囲で自由に回転することができる。この隙間を「バックラッシ」（又は「バックラッシュ」）と呼ぶ。

50

【 0 0 0 3 】

例えば、モータと主軸等の被駆動軸がギアで結合されている場合、ギアのバックラッシが存在することにより、モータの反転時には、被駆動軸の反転がモータの反転に対して遅れてしまうこととなる。この被駆動軸の反転遅れを補正するために、モータへの位置指令にバックラッシ補正量を加算する方法が、従来存在する。これを「バックラッシ補正」と呼ぶ。

【 0 0 0 4 】

一方で、例えば被駆動軸における静摩擦等の影響により、被駆動軸の反転がモータの反転に対して遅れてしまうことがある。この被駆動軸の反転遅れを補正するために、モータ反転時に、位置指令ではなく、速度指令又はトルク指令に補正を加算する方法が、従来存在する。これを「バックラッシ加速補正」と呼ぶ。

10

【 0 0 0 5 】

バックラッシ補正制御及びバックラッシ加速補正制御を用いたサーボ制御装置としては、例えば、特許文献 1 に記載されたサーボ制御装置がある。特許文献 1 に記載された補正パラメータ調整装置としてのサーボ制御装置は、実際の使用状態にある機械の運動軌跡をジグや工具を外さずに測定し、誤差要因を分離して補正パラメータの適切な設定を可能とするため、機械の加速度を測定するための加速度計と、機械運動を解析する機械運動解析部と、機械の運動精度を向上するための、バックラッシ補正パラメータ等を決定する補正パラメータ計算部とを備える。

【 先行技術文献 】

20

【 特許文献 】

【 0 0 0 6 】

【 特許文献 1 】 特開 2 0 1 1 - 2 2 1 6 1 2 号公報

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 7 】

上記のように、サーボ制御装置においては、機械の摩擦、ガタ、ロストモーション等の非線形特性の影響により、位置指令値や速度指令値が変化したときに位置偏差が発生する場合がある。このような場合に、バックラッシ補正やバックラッシ加速補正を時間や距離に応じた複雑な形状（以下、「複雑な形状」ともいう）にすることで、位置偏差や速度偏差を低減し、位置指令や速度指令に対する追従性を向上できる。しかし、バックラッシ補正やバックラッシ補正を複雑な形状にするほど、バックラッシ補正やバックラッシ加速補正の際に用いられるパラメータ（「係数」ともいう）の調整は難しくなる。

30

【 0 0 0 8 】

本発明は、バックラッシ補正やバックラッシ加速補正を用いたサーボ制御装置において、位置偏差や速度偏差を低減して、位置指令に対する追従性を向上させるためにバックラッシ補正やバックラッシ加速補正を複雑な形状にした場合に、複雑な形状のバックラッシ補正やバックラッシ加速補正のパラメータ（係数）の煩雑な調整を回避できる強化学習を行う機械学習装置、この機械学習装置を含むサーボ制御装置、サーボ制御システム、及び機械学習方法を提供することを目的とする。

40

【 課題を解決するための手段 】

【 0 0 0 9 】

（ 1 ） 本発明に係る機械学習装置（例えば、後述の「機械学習装置 2 0 0 」）は、位置指令又は位置偏差に対するバックラッシ補正量を作成するバックラッシ補正手段（例えば、後述の「バックラッシ補正部 1 0 9 」）及び／又は、速度指令に対するバックラッシ加速補正量を作成するバックラッシ加速補正手段（例えば、後述の「バックラッシ加速補正部 1 1 0 」）を備える、サーボ制御装置に対して、前記バックラッシ補正手段におけるバックラッシ補正パラメータ及び／又は前記バックラッシ加速補正手段におけるバックラッシ加速補正パラメータを対象とする強化学習を行う機械学習装置であって、前記バックラッシ補正パラメータ及び／又は前記バックラッシ加速補正パラメータの調整情報を含む

50

行動情報を前記バックラッシュ補正手段及び／又は前記バックラッシュ加速補正手段に対して出力する行動情報出力手段（例えば、後述の「行動情報出力部２０３」）と、前記行動情報に基づいて、所定の加工プログラムを前記サーボ制御装置に実行させた場合における、前記位置指令と位置フィードバックから得られる位置偏差と、前記バックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正パラメータと、を含む状態情報を前記サーボ制御装置から取得する状態情報取得手段（例えば、後述の「状態情報取得部２０１」）と、前記状態情報に含まれる前記位置偏差に基づいて、強化学習における報酬の値を出力する報酬出力手段（例えば、後述の「報酬出力部２０２１」）と、前記報酬の値と、前記状態情報と、前記行動情報とに基づいて行動価値関数を更新する価値関数更新手段（例えば、後述の「価値関数更新部２０２２」）と、を備える。

10

【００１０】

（２） 上記（１）の機械学習装置において、前記バックラッシュ補正手段におけるバックラッシュ補正パラメータを対象として強化学習を行った後に、前記バックラッシュ加速補正手段における前記バックラッシュ加速補正パラメータを対象として強化学習を行うようにしてもよい。

【００１１】

（３） 上記（１）又は（２）の機械学習装置において、前記報酬出力手段は、前記位置偏差の絶対値に基づいて前記報酬の値を出力してもよい。

【００１２】

（４） 上記（１）から（３）の何れかの機械学習装置において、前記価値関数更新手段により更新された行動価値関数に基づいて、前記バックラッシュ補正手段におけるバックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正手段における前記バックラッシュ加速補正パラメータを生成して出力する最適化行動情報出力手段（例えば、後述の「最適化行動情報出力部２０５」）を備えてもよい。

20

【００１３】

（５） 本発明に係るサーボ制御装置は、上記（１）から（４）の何れかの機械学習装置を含むサーボ制御装置である。

【００１４】

（６） 本発明に係るサーボ制御システムは、上記（１）から（４）の何れかの機械学習装置と、前記サーボ制御装置と、を備えたサーボ制御システムである。

30

【００１５】

（７） 本発明に係る機械学習方法は、位置指令又は位置偏差に対するバックラッシュ補正量を作成するバックラッシュ補正手段と、速度指令に対するバックラッシュ加速補正量を作成するバックラッシュ加速補正手段とを備える、サーボ制御装置に対して前記バックラッシュ補正手段におけるバックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正手段におけるバックラッシュ加速補正パラメータを対象とする強化学習を行う機械学習装置の機械学習方法であって、前記バックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正パラメータの調整情報を含む行動情報を前記バックラッシュ補正手段及び／又は前記バックラッシュ加速補正手段に対して出力する行動情報出力ステップと、前記行動情報に基づいて、所定の加工プログラムを前記サーボ制御装置に実行させた場合における、前記位置指令と位置フィードバックから得られる位置偏差と、前記バックラッシュ補正パラメータ及び／又は前記バックラッシュ加速補正パラメータと、を含む状態情報を前記サーボ制御装置から取得する状態情報取得ステップと、前記状態情報に含まれる前記位置偏差に基づいて、強化学習における報酬の値を出力する報酬出力ステップと、前記報酬の値と、前記状態情報と、前記行動情報とに基づいて行動価値関数を更新する価値関数更新ステップと、を備える。

40

【発明の効果】**【００１６】**

本発明によれば、バックラッシュ補正やバックラッシュ加速補正を用いたサーボ制御装置において、位置偏差や速度偏差を低減して、位置指令に対する追従性を向上させるためにバ

50

ックラッシュ補正やバックラッシュ加速補正を複雑な形状にした場合に、複雑な形状のバックラッシュ補正やバックラッシュ加速補正の係数の煩雑な調整を回避できる強化学習を行うことができる。

【図面の簡単な説明】

【0017】

【図1】本発明の第1の実施形態のサーボ制御システムを示すブロック図である。

【図2】サーボ制御装置100と機械学習装置200がネットワークで接続された構成例を示すブロック図である。

【図3】制御対象300の一例を示すブロック図である。

【図4A】加工形状が円形の場合のサーボモータの動作を説明するための図である。

10

【図4B】加工形状が四角の場合のサーボモータの動作を説明するための図である。

【図4C】加工形状が角R付き四角の場合のサーボモータの動作を説明するための図である。

【図4D】加工形状が円形で、位置A1で、テーブルをY軸方向に移動するサーボモータの回転方向が反転しようとした時の象限突起を示す説明図である。

【図5】第1の実施形態の機械学習装置200を示すブロック図である。

【図6】台形型加速方式に基づくバックラッシュ加速補正の概要を示す図である。

【図7A】バックラッシュ加速補正の際の加速量V0の大きさと加速度の平方根との関係を示す図である。

【図7B】バックラッシュ加速補正の際の加速量V0の大きさと加速度の平方根との関係を示す図である。

20

【図7C】バックラッシュ加速補正の際の加速量V0の大きさと加速度の平方根との関係を示す図である。

【図8】第1の強化学習時におけるサーボ制御装置100と機械学習装置200の構成例を示すブロック図である。

【図9】機械学習装置200の動作を説明するフローチャートである。

【図10】機械学習装置200の最適化行動情報出力部205の動作を説明するフローチャートである。

【図11】第2の強化学習時におけるサーボ制御装置100と機械学習装置200の構成例を示すブロック図である。

30

【図12】機械学習終了後の運用時におけるサーボ制御装置100の構成例を示すブロック図である。

【発明を実施するための形態】

【0018】

以下、本発明の実施形態について図面を用いて詳細に説明する。

(第1の実施形態)

図1は発明の第1の実施形態のサーボ制御システムを示すブロック図である。サーボ制御システム10は、図1に示すように、例えばn台のサーボ制御装置100-1~100-n、n台の機械学習装置200-1~200-n、及びネットワーク400を備えている。なお、nは任意の自然数である。

40

【0019】

ここで、サーボ制御装置100-1と機械学習装置200-1とは1対1の組とされて、通信可能に接続されている。サーボ制御装置100-2~100-nと機械学習装置200-2~200-nについてもサーボ制御装置100-1と機械学習装置200-1と同様に接続される。図1では、サーボ制御装置100-1~100-nと機械学習装置200-1~200-nとのn個の組は、ネットワーク400を介して接続されているが、サーボ制御装置100-1~100-nと機械学習装置200-1~200-nとのn個の組は、それぞれの組のサーボ制御装置と機械学習装置とが接続インタフェースを介して直接接続されてもよい。これらサーボ制御装置100-1~100-nと機械学習装置200-1~200-nとのn個の組は、例えば同じ工場に複数組設置されていてもよく、

50

それぞれ異なる工場に設置されていてもよい。

なお、ネットワーク 400 は、例えば、工場内に構築された LAN (Local Area Network) や、インターネット、公衆電話網、或いは、これらの組み合わせである。ネットワーク 400 における具体的な通信方式や、有線接続及び無線接続の何れであるか等については、特に限定されない。

【0020】

図 2 は本発明の第 1 の実施形態のサーボ制御システムのサーボ制御装置と機械学習装置との組、及び制御対象を示すブロック図である。図 2 のサーボ制御装置 100 と、機械学習装置 200 は例えば、図 1 に示すサーボ制御装置 100 - 1 と機械学習装置 200 - 1 とに対応している。

10

制御対象 300 は例えばサーボモータ、サーボモータを含む工作機械、ロボット、産業機械等である。サーボ制御装置 100 は工作機械、ロボット、産業機械等の一部として設けられてもよい。

【0021】

最初に、サーボ制御装置 100 について説明する。

サーボ制御装置 100 は、位置指令作成部 101、減算器 102、加算器 103、位置制御部 104、減算器 105、加算器 106、速度制御部 107、積分器 108、バックラッシュ補正部 109、及びバックラッシュ加速補正部 110 を備えている。

【0022】

位置指令作成部 101 は位置指令値を作成し、作成した位置指令値を減算器 102 に出力する。減算器 102 は位置指令値とフィードバックされた検出位置との差を求め、その差を位置偏差値として加算器 103 とに出力する。なお、機械学習時には当該位置偏差値は後述の機械学習装置 200 に対しても送信される。

20

【0023】

位置指令作成部 101 は、制御対象 300 を動作させる任意の加工プログラムに基づいて位置指令値を作成する。制御対象 300 は、例えばサーボモータを含む工作機械で、被加工物 (ワーク) を搭載するテーブルを X 軸方向及び Y 軸方向に移動させて加工する場合には、X 軸方向及び Y 軸方向に対してそれぞれ図 2 に示すサーボ制御装置が設けられる。テーブルを 3 軸以上の方向に移動させる場合には、それぞれの軸方向に対してサーボ制御装置が設けられる。

30

位置指令作成部 101 は、加工プログラムにより指定される加工形状となるように、送り速度を設定して位置指令値を作成する。

【0024】

バックラッシュ補正部 109 は、バックラッシュ補正パラメータに基づいて作成されるバックラッシュ補正量を加算器 103 に出力する。

【0025】

加算器 103 は、減算器 102 から入力される位置偏差に、バックラッシュ補正部 109 から入力されたバックラッシュ補正量を加算し、バックラッシュ補正量を加算した位置偏差を、位置制御部 104 に出力する。なお、加算器 103 は、位置偏差ではなく、位置指令にバックラッシュ補正量を加算することも可能である。

40

【0026】

位置制御部 104 は、位置偏差にポジションゲイン K_p を乗じた値を、速度指令値として減算器 105 に出力する。減算器 105 は速度指令値とフィードバックされた速度検出値との差を求め、その差を速度偏差として加算器 106 に出力する。

【0027】

バックラッシュ加速補正部 110 は、バックラッシュ加速補正パラメータに基づいて作成されるバックラッシュ加速補正量を加算器 106 に出力する。

【0028】

加算器 106 は、減算器 105 から入力される速度偏差に、バックラッシュ加速補正部 110 から入力されたバックラッシュ加速補正量を加算し、バックラッシュ加速補正量を加算し

50

た速度偏差を、速度制御部 107 に出力する。

【0029】

速度制御部 107 は、速度偏差に積分ゲイン $K_1 v$ を乗じて積分した値と、速度偏差に比例ゲイン $K_2 v$ を乗じた値とを加算して、トルク指令値として制御対象 300 に出力する。

【0030】

制御対象 300 は速度検出値を出力し、速度検出値は速度フィードバックとして減算器 105 に入力される。速度検出値は積分器 108 で積分されて位置検出値となり、位置検出値は位置フィードバックとして減算器 102 に入力される。

以上のように、サーボ制御装置 100 は構成される。

10

【0031】

次に、サーボ制御装置 100 により制御される制御対象 300 について説明する。

図 3 は制御対象 300 の一例となる、サーボモータを含む工作機械を示すブロック図である。

サーボ制御装置 100 は、サーボモータ 302 で連結機構 303 を介してテーブル 304 を移動させ、テーブル 304 の上に搭載された被加工物（ワーク）を加工する。連結機構 303 は、サーボモータ 302 に連結されたカップリング 3031 と、カップリング 3031 に固定されるボールねじ 3033 とを有し、ボールねじ 3033 にナット 3032 が螺合されている。サーボモータ 302 の回転駆動によって、ボールねじ 3033 に螺着されたナット 3032 がボールねじ 3033 の軸方向に移動する。

20

【0032】

サーボモータ 302 の回転角度位置は、サーボモータ 302 に関連付けられた、位置検出部となるロータリーエンコーダ 301 によって検出され、検出された信号は速度フィードバックとして利用される。検出された信号は積分することで位置フィードバックとして利用される。なお、ボールねじ 3033 の端部に取り付けられ、ボールねじ 3033 の移動距離を検出するリニアスケール 305 の出力を位置フィードバックとして用いてもよい。また、加速度センサを用いて位置フィードバックを生成してもよい。

【0033】

機械学習装置 200 は、予め用意される所定の加工プログラム（以下、「学習時の加工プログラム」ともいう）を実行することで、バックラッシュ補正部 109 がバックラッシュ補正量を作成する際に用いるバックラッシュ補正パラメータ、及びバックラッシュ加速補正部 110 がバックラッシュ加速補正量を作成する際に用いるバックラッシュ加速補正パラメータを対象として強化学習を行う。ここで、学習時の加工プログラムにより指定される加工形状は、例えば、円、四角、及び角 R 付き四角（a square with quarter arc）等である。

30

【0034】

学習時の加工プログラムについて説明する。

図 4 A は、学習時の加工プログラムにより指定される加工形状が円形の場合のサーボモータの動作を説明するための図である。図 4 B は、学習時の加工プログラムにより指定される加工形状が四角の場合のサーボモータの動作を説明するための図である。図 4 C は、学習時の加工プログラムにより指定される加工形状が角 R 付き四角の場合のサーボモータの動作を説明するための図である。図 4 A ~ 図 4 C において、被加工物（ワーク）が時計まわりに加工されるようにテーブルが移動する。

40

【0035】

加工形状が円形の場合は、図 4 A に示すように、位置 A 1 と位置 A 3 で、テーブルを Y 軸方向に移動するサーボモータは回転方向が反転し、テーブルは Y 軸方向に直線反転するように移動する。位置 A 2 と位置 A 4 で、テーブルを X 軸方向に移動するサーボモータは、回転方向が反転し、テーブルは X 軸方向に直線反転するように移動する。

【0036】

加工形状が四角の場合は、図 4 B に示すように、位置 B 1 で、テーブルを Y 軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルを X 軸方向に移動するサーボ

50

モータは、回転動作から停止に移り、テーブルはX軸方向の直線動作からY軸方向の直線動作に移る。同様に、位置B2で、テーブルをX軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルをY軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルはY軸方向の直線動作からX軸方向の直線動作に移る。位置B3で、テーブルをY軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルをX軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルはX軸方向の直線動作からY軸方向の直線動作に移る。位置B4で、テーブルをX軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルをY軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルはY軸方向の直線動作からX軸方向の直線動作に移る。

【0037】

加工形状が角R付き四角の場合は、図4Cに示すように、位置C1で、テーブルをY軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルはX軸方向の直線動作から円弧動作に移る。位置C2で、テーブルをX軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルは円弧動作からY軸方向の直線動作に移る。同様に、位置C3で、テーブルをX軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルはY軸方向の直線動作から円弧動作に移る。位置C4で、テーブルをY軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルは円弧動作からX軸方向の直線動作に移る。位置C5で、テーブルをY軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルはX軸方向の直線動作から円弧動作に移る。位置C6で、テーブルをX軸方向に移動するサーボモータは回転動作から停止に移り、テーブルは円弧動作からY軸方向の直線動作に移る。位置C7で、テーブルをX軸方向に移動するサーボモータは、停止から回転動作に移り、テーブルはY軸方向の直線動作から円弧動作に移る。位置C8で、テーブルをY軸方向に移動するサーボモータは、回転動作から停止に移り、テーブルは円弧動作からX軸方向の直線動作に移る。

【0038】

図4Dは、加工形状が円形で、位置A1で、テーブルをY軸方向に移動するサーボモータの回転方向が反転しようとした時に、機械要素間の抵触面におけるガタや摩擦により生ずる象限突起の生成を示す説明図である。

図4Dに示すように、テーブルをY軸方向に移動するサーボモータの回転方向が位置A1で反転しようとした時に、機械要素間の抵触面にガタが存在することにより、軌跡の回転中心がずれてしまう。また、摩擦によって反転遅れが生じることにより、半径方向に軌跡誤差が拡大し、軌跡誤差の拡大表示を行うと軌跡誤差が突起として見える。回転中心のずれに対しては、バックラッシュ補正が必要であり、突起の解消に対しては、バックラッシュ加速補正が必要である。

以上のように、学習時の加工プログラムにより指定される加工形状により、回転方向が反転したり、回転状態から停止したりする場合に生ずる惰走（惰性で動作する）を評価し、位置偏差に対する影響を調べることができる。

【0039】

学習時の加工プログラムを実行することで、サーボ制御装置100の位置指令作成部101は順次、円、四角、角R付き四角（a square with quarter arc）の加工形状となるように位置指令値を出力する。また、円、四角、角R付き四角（a square with quarter arc）の加工形状毎に、送り速度を変更し、複数の送り速度についての影響も学習できるようにする。送り速度は、加工形状の図形の移動途中、例えば四角の加工形状にテーブルを移動させているときに、角を過ぎたときに送り速度を変更してもよい。

【0040】

<強化学習について>

機械学習装置200に含まれる各機能ブロックの説明に先立って、まず強化学習の基本的な仕組みについて説明する。エージェント（本実施形態における機械学習装置200に相当）は、環境の状態を観測し、ある行動を選択し、当該行動に基づいて環境が変化する。環境の変化に伴って、何らかの報酬が与えられ、エージェントはより良い行動の選択（

10

20

30

40

50

意思決定)を学習する。

教師あり学習が、完全な正解を示すのに対して、強化学習における報酬は、環境の一部の変化に基づく断片的な値であることが多い。このため、エージェントは、将来にわたっての報酬の合計を最大にするよう行動を選択するように学習する。

【0041】

このように、強化学習では、行動を学習することにより、環境に行動が与える相互作用を踏まえて適切な行動を学習、すなわち将来的に得られる報酬を最大にするための学習する方法を学ぶ。これは、本実施形態において、例えば、位置偏差を低減するための行動情報を選択するという、未来に影響をおよぼすような行動を獲得できることを表している。

【0042】

ここで、強化学習としては、任意の学習方法を用いることができるが、以下の説明では、或る環境の状態 s の下で、行動 a を選択する価値 $Q(s, a)$ を学習する方法である Q 学習 (Q -learning) を用いる場合を例にとって説明をする。

Q 学習では、或る状態 s のとき、取り得る行動 a のなかから、価値 $Q(s, a)$ の最も高い行動 a を最適な行動として選択することを目的とする。

【0043】

しかしながら、 Q 学習を最初に開始する時点では、状態 s と行動 a との組合せについて、価値 $Q(s, a)$ の正しい値は全く分かっていない。そこで、エージェントは、或る状態 s の下で様々な行動 a を選択し、その時の行動 a に対して、与えられる報酬に基づいて、より良い行動の選択をすることにより、正しい価値 $Q(s, a)$ を学習していく。

【0044】

また、将来にわたって得られる報酬の合計を最大化したいので、最終的に $Q(s, a) = E[\sum(\gamma^t) r_t]$ となるようにすることを目指す。ここで $E[\]$ は期待値を表し、 t は時刻、 γ は後述する割引率と呼ばれるパラメータ、 r_t は時刻 t における報酬、 $\sum$ は時刻 t による合計である。この式における期待値は、最適な行動に従って状態変化した場合の期待値である。しかし Q 学習の過程において最適な行動が何であるのかは不明であるので、様々な行動を行うことにより、探索しながら強化学習をする。このような価値 $Q(s, a)$ の更新式は、例えば、次の数式 1 (以下に数 1 として示す) により表すことができる。

【0045】

【数 1】

$$Q(s_{t+1}, a_{t+1}) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$$

【0046】

上記の数式 1 において、 s_t は、時刻 t における環境の状態を表し、 a_t は、時刻 t における行動を表す。行動 a_t により、状態は s_{t+1} に変化する。 r_{t+1} は、その状態の変化により得られる報酬を表している。また、 $\max$ の付いた項は、状態 s_{t+1} の下で、その時に分かっている最も Q 値の高い行動 a を選択した場合の Q 値に γ を乗じたものになる。ここで、 γ は、 $0 < \gamma < 1$ のパラメータで、割引率と呼ばれる。また、 α は、学習係数で、 $0 < \alpha < 1$ の範囲とする。

【0047】

上述した数式 1 は、試行 a_t の結果、返ってきた報酬 r_{t+1} を元に、状態 s_t における行動 a_t の価値 $Q(s_t, a_t)$ を更新する方法を表している。

この更新式は、状態 s_t における行動 a_t の価値 $Q(s_t, a_t)$ よりも、行動 a_t による次の状態 s_{t+1} における最良の行動の価値 $\max_a Q(s_{t+1}, a)$ の方が大きければ、 $Q(s_t, a_t)$ を大きくし、逆に小さければ、 $Q(s_t, a_t)$ を小さくすることを示している。つまり、或る状態における或る行動の価値を、それによる次の状態における最良の行動の価値に近づける。ただし、その差は、割引率 γ と報酬 r_{t+1} のあ

10

20

30

40

50

り方により変わってくるが、基本的には、ある状態における最良の行動の価値が、それに至る一つ前の状態における行動の価値に伝播していく仕組みになっている。

【 0 0 4 8 】

ここで、Q 学習では、すべての状態行動ペア (s, a) についての $Q(s, a)$ のテーブルを作成して、学習を行う方法がある。しかし、すべての状態行動ペアの $Q(s, a)$ の値を求めるには状態数が多すぎて、Q 学習が収束するのに多くの時間を要してしまう場合がある。

【 0 0 4 9 】

そこで、公知の DQN (Deep Q - Network) と呼ばれる技術を利用するようにしてもよい。具体的には、行動価値関数 Q を適当なニューラルネットワークを用いて構成し、ニューラルネットワークのパラメータを調整することにより、価値関数 Q を適当なニューラルネットワークで近似することにより価値 $Q(s, a)$ の値を算出するようにしてもよい。DQN を利用することにより、Q 学習が収束するのに要する時間を短くすることが可能となる。なお、DQN については、例えば、以下の非特許文献に詳細な記載がある。

【 0 0 5 0 】

< 非特許文献 >

「Human-level control through deep reinforcement learning」、Volodymyr Mnih 著 [online]、[平成 29 年 1 月 17 日検索]、インターネット URL : <http://files.davidqu.com/research/nature14236.pdf>

【 0 0 5 1 】

以上説明をした Q 学習を機械学習装置 200 が行う。

なお、本実施形態では、バックラッシュ補正部 109 が作成するバックラッシュ補正量の算出に用いるバックラッシュ補正パラメータ、及びバックラッシュ加速補正部 110 がバックラッシュ加速補正量を作成する際に用いるバックラッシュ加速補正パラメータについて同時に学習するのではなく、先ずバックラッシュ補正パラメータを対象とする強化学習 (「第 1 の強化学習」ともいう) を行う。第 1 の強化学習により、最適なバックラッシュ補正パラメータを算出すると、当該バックラッシュ補正パラメータをバックラッシュ補正部 109 に設定したうえで、バックラッシュ加速補正部 110 がバックラッシュ加速補正量を作成する際に用いるバックラッシュ加速補正パラメータを対象とする強化学習 (「第 2 の強化学習」ともいう) を行う。

こうすることで、強化学習を効率的に行うことができる。

【 0 0 5 2 】

< 第 1 の強化学習 >

まず、第 1 の強化学習について説明する。第 1 の強化学習において、機械学習装置 200 は、サーボ制御装置 100 におけるバックラッシュ補正部 109 が作成するバックラッシュ補正量の算出に用いるバックラッシュ補正パラメータ、並びに学習時の位置指令に応じた制御を実行することで取得されるサーボ制御装置 100 の位置偏差情報を含むサーボ状態を状態 s として、当該状態 s に係る、バックラッシュ補正パラメータの調整を行動 a として選択する価値 Q を学習する。なお、第 1 の強化学習時においては、バックラッシュ加速補正を省略してもよい。

【 0 0 5 3 】

バックラッシュ補正のための学習においては、バックラッシュ補正パラメータ P_0 の調整 (修正) を、行動 a として選択する価値 Q を学習する。

【 0 0 5 4 】

機械学習装置 200 は、バックラッシュ補正パラメータ P_0 に基づいて、学習時の加工プログラムをサーボ制御装置 100 に実行させた場合の位置指令と位置フィードバック値から得られる位置偏差値の集合と、バックラッシュ補正パラメータ P_0 と、を含む状態情報 s を観測して、次の行動 a (バックラッシュ補正パラメータ P_0 の調整) を決定する。

機械学習装置 200 は、行動 a をするたびに報酬が返ってくる。機械学習装置 200 は

10

20

30

40

50

、例えば、将来にわたっての報酬の合計が最大になるように最適な行動 a を試行錯誤的に探索する。そうすることで、機械学習装置 200 は、バックラッシュ補正パラメータ P_0 に基づいて、学習時の加工プログラムをサーボ制御装置 100 に実行させることで得られるサーボ制御装置 100 の位置偏差情報を含むサーボ状態を含む状態情報 s に対して、最適な行動 a 、すなわち、バックラッシュ補正パラメータ P_0 を選択することが可能となる。

【0055】

これにより、機械学習装置 200 により学習された価値関数 Q に基づいて、或る状態 s に係るバックラッシュ補正パラメータ P_0 に対して適用される行動 a のうち、 Q の値が最大となるような行動 a を選択することで、位置偏差が最小になるような行動 a （すなわち、バックラッシュ補正部 109 の最適なバックラッシュ補正パラメータ P_0 を選択すること）が

10

【0056】

すなわち、機械学習装置 200 により学習された価値関数 Q に基づいて、或る状態 s に係るバックラッシュ補正パラメータ P_0 に対して適用される行動 a のうち、 Q の値が最大となるような行動 a を選択することで、学習時の加工プログラムを実行することで取得される位置偏差が最小になるような行動 a （すなわち、バックラッシュ補正パラメータ P_0 ）を選択することが可能となる。

【0057】

なお、本実施形態における強化学習は、バックラッシュ補正パラメータ P_0 を対象としたものとして説明したが、バックラッシュ補正パラメータ P_0 に限定されない。

20

例えば、サーボモータの回転方向が反転した後、反転後の移動量が固定値 L_0 に達するまでは補正を行わず、移動量が固定値 L_0 に達した時点において、補正量がバックラッシュ補正パラメータ P_0 のバックラッシュ補正を実行するケースについて強化学習を行ってもよい。すなわち、バックラッシュ補正パラメータとして P_0 及び L_0 を対象として強化学習を行ってもよい。

この場合、行動 a は、バックラッシュ補正パラメータ P_0 及び L_0 の調整となる。

また、サーボモータの回転方向が右方向（順方向）か、左方向（逆方向）かによって、補正量及び／又は移動量として、右方向（順方向）のバックラッシュ補正パラメータ及び左方向（逆方向）のバックラッシュ補正パラメータによるバックラッシュ補正を実行するケースについて強化学習を行ってもよい。すなわち、右方向（順方向）のバックラッシュ補正パラメータ及び左方向（逆方向）のバックラッシュ補正パラメータを対象として強化学習を行ってもよい。

30

【0058】

< 機械学習装置 200 の機能ブロック >

図 5 は第 1 の強化学習を行う場合の機械学習装置 200 を示すブロック図である。

上述した第 1 の強化学習を行うために、図 5 に示すように、機械学習装置 200 は、状態情報取得部 201、学習部 202、行動情報出力部 203、価値関数記憶部 204、及び最適化行動情報出力部 205 を備える。学習部 202 は報酬出力部 2021、価値関数更新部 2022、及び行動情報生成部 2023 を備える。

【0059】

40

なお、状態情報取得部 201、学習部 202、行動情報出力部 203、価値関数記憶部 204、最適化行動情報出力部 205、報酬出力部 2021、価値関数更新部 2022、行動情報生成部 2023 は、上記のバックラッシュ補正パラメータ P_0 を対象とした強化学習のために動作する。

【0060】

状態情報取得部 201 は、サーボ制御装置 100 におけるバックラッシュ補正パラメータ P_0 に基づいて、学習時の加工プログラムを実行することで取得されるサーボ制御装置 100 の位置偏差情報を含む指令、フィードバック等のサーボ状態を含む状態 s を、サーボ制御装置 100 から取得する。この状態情報 s は、 Q 学習における、環境状態 s に相当する。

50

状態情報取得部 201 は、取得した状態情報 s を学習部 202 に対して出力する。

なお、最初に Q 学習を開始する時点での、バックラッシュ補正パラメータ P_0 は、予めユーザが生成するようにする。本実施形態では、ユーザが作成したバックラッシュ補正パラメータ P_0 の初期設定値を、強化学習により最適なものに調整する。

【0061】

学習部 202 は、或る環境状態 s の下で、ある行動 a を選択する場合の価値 $Q(s, a)$ を学習する部分である。具体的には、学習部 202 は、報酬出力部 2021、価値関数更新部 2022 及び行動情報生成部 2023 を備える。

【0062】

報酬出力部 2021 は、或る状態 s の下で、行動 a を選択した場合の報酬を算出する部分である。ここで、状態 s における状態変数である位置偏差値の集合（位置偏差集合）を $PD(s)$ 、行動情報 a （バックラッシュ補正パラメータ P_0 の修正）により状態 s から変化した状態情報 s' に係る状態変数である位置偏差集合を $PD(s')$ で示す。また、状態 s における位置偏差の値を、予め設定された評価関数 $f(PD(s))$ に基づいて算出される値とする。

評価関数 f としては、例えば、

位置偏差の絶対値の積算値を算出する関数

$$\int |e| dt$$

位置偏差の絶対値に時間の重み付けをして積算値を算出する関数

$$\int t |e| dt$$

位置偏差の絶対値の $2n$ (n は自然数) 乗の積算値を算出する関数、

$$\int e^{2n} dt \quad (n \text{ は自然数})$$

位置偏差の絶対値の最大値を算出する関数

$$\text{Max} \{ |e| \}$$

等を適用することができる。

【0063】

このとき、行動情報 a により修正された状態情報 s' に係る修正後のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s'))$ が、行動情報 a により修正される前の状態情報 s に係る修正前のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s))$ よりも大きくなった場合に、報酬出力部 2021 は、報酬の値を負の値とする。

【0064】

一方で、行動情報 a により修正された状態情報 s' に係る修正後のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s'))$ が、行動情報 a により修正される前の状態情報 s に係る修正前のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s))$ よりも小さくなった場合に、報酬の値を正の値とする。

なお、行動情報 a により修正された状態情報 s' に係る修正後のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s'))$ が、行動情報 a により修正される前の状態情報 s に係る修正前のバックラッシュ補正部 109 に基づいて動作したサーボ制御装置 100 の位置偏差の値 $f(PD(s))$ と等しい場合は、報酬出力部 2021 は、報酬の値をゼロとする。

【0065】

また、行動 a を実行後の状態 s' の位置偏差の値 $f(PD(s'))$ が、前の状態 s における位置偏差の値 $f(PD(s))$ より大きくなった場合の負の値としては、その比率に応じて負の値を大きくするようにしてもよい。つまり位置偏差の値が大きくなった度合いに応じて負の値が大きくなるようにするとよい。逆に、行動 a を実行後の状態 s' の位置偏差の値 $f(PD(s'))$ が、前の状態 s における位置偏差の値 $f(PD(s))$ より小さくなった場合の正の値としては、比率に応じて正の値を大きくするようにしてもよい。つまり位置偏差の値が小さくなった度合いに応じて正の値が大きくなるようにすると

10

20

30

40

50

よい。

また、各係数の値をランダムに設定してもよい。

【0066】

価値関数更新部2022は、状態 s と、行動 a と、行動 a を状態 s に適用した場合の状態 s' と、上記のようにして算出された報酬の値と、に基づいてQ学習を行うことにより、価値関数記憶部204が記憶する行動価値関数 Q を更新する。

行動価値関数 Q の更新は、オンライン学習で行ってもよく、バッチ学習で行ってもよく、ミニバッチ学習で行ってもよい。

オンライン学習とは、或る行動 a を現在の状態 s に適用することにより、状態 s が新たな状態 s' に遷移する都度、即座に行動価値関数 Q の更新を行うという学習方法である。また、バッチ学習とは、或る行動 a を現在の状態 s に適用することにより、状態 s が新たな状態 s' に遷移することを繰り返すことにより、学習用のデータを収集し、収集したすべての学習用データを用いて、行動価値関数 Q の更新を行うという学習方法である。更に、ミニバッチ学習とは、オンライン学習と、バッチ学習の中間的な、ある程度学習用データが溜まるたびに行動価値関数 Q の更新を行うという学習方法である。

【0067】

行動情報生成部2023は、現在の状態 s に対して、Q学習の過程における行動 a を選択する。行動情報生成部2023は、Q学習の過程において、サーボ制御装置100のバックラッシュ補正パラメータ P_0 を修正する動作（Q学習における行動 a に相当）を行わせるために、行動情報 a を生成して、生成した行動情報 a を行動情報出力部203に対して出力する。より具体的には、行動情報生成部2023は、例えば、バックラッシュ補正部109のバックラッシュ補正パラメータ P_0 をインクレメンタル（例えば0.01程度）に加算又は減算させる。

【0068】

そして、行動情報生成部2023は、バックラッシュ補正パラメータ P_0 の増加又は減少を適用して、状態 s' に遷移して、プラスの報酬（正の値の報酬）が返った場合、次の行動 a' としては、バックラッシュ補正パラメータ P_0 に対して、前回のアクションと同様にインクレメンタルに加算又は減算させる等、位置偏差の値がより小さくなるような行動 a' を選択する方策を取るようにしてもよい。

【0069】

また、逆に、マイナスの報酬（負の値の報酬）が返った場合、行動情報生成部2023は、次の行動 a' としては、例えば、バックラッシュ補正部109のバックラッシュ補正パラメータ P_0 に対して、前回のアクションとは逆にインクレメンタルに減算又は加算させる等、位置偏差が前回の値よりも小さくなるような行動 a' を選択する方策を取るようにしてもよい。

【0070】

また、行動情報生成部2023は、現在の推定される行動 a の価値の中で、最も価値 $Q(s, a)$ の高い行動 a' を選択するグリーディ法や、ある小さな確率 ϵ でランダムに行動 a' 選択し、それ以外では最も価値 $Q(s, a)$ の高い行動 a' を選択する ϵ グリーディ法といった公知の方法により、行動 a' を選択する方策を取るようにしてもよい。

【0071】

行動情報出力部203は、学習部202から出力される行動情報 a をサーボ制御装置100に対して送信する部分である。サーボ制御装置100は上述したように、この行動情報に基づいて、現在の状態 s 、すなわち現在設定されているバックラッシュ補正パラメータ P_0 を調整することで、次の状態 s' （すなわち修正された、バックラッシュ補正パラメータ P_0 ）に遷移する。

【0072】

価値関数記憶部204は、行動価値関数 Q を記憶する記憶装置である。行動価値関数 Q は、例えば状態 s 、行動 a 毎にテーブル（以下、行動価値テーブルと呼ぶ）として格納してもよい。価値関数記憶部204に記憶された行動価値関数 Q は、価値関数更新部202

10

20

30

40

50

2により更新される。また、価値関数記憶部204に記憶された行動価値関数Qは、他の機械学習装置200との間で共有されるようにしてもよい。行動価値関数Qを複数の機械学習装置200で共有するようにすれば、各機械学習装置200にて分散して強化学習を行うことが可能となるので、強化学習の効率を向上させることが可能となる。

【0073】

最適化行動情報出力部205は、価値関数更新部2022がQ学習を行うことにより更新した行動価値関数Qに基づいて、価値 $Q(s, a)$ が最大となる動作をバックラッシュ補正部109に行わせるための行動情報a(以下、「最適化行動情報」と呼ぶ)を生成する。

より具体的には、最適化行動情報出力部205は、価値関数記憶部204が記憶している行動価値関数Qを取得する。この行動価値関数Qは、上述したように価値関数更新部2022がQ学習を行うことにより更新したものである。そして、最適化行動情報出力部205は、行動価値関数Qに基づいて、行動情報を生成し、生成した行動情報をサーボ制御装置100(バックラッシュ補正部109)に対して出力する。この最適化行動情報には、行動情報出力部203AがQ学習の過程において出力する行動情報と同様に、バックラッシュ補正部109の係数P0を修正する情報が含まれる。

【0074】

サーボ制御装置100では、この行動情報に基づいてバックラッシュ補正部109に修正されたバックラッシュ補正パラメータ P_0 が設定され、位置偏差の値を低減するように動作することができる。

以上のように、本発明に係る機械学習装置200を利用することで、サーボ制御装置100のバックラッシュ補正部109に設定されたバックラッシュ補正パラメータ P_0 の最適化を容易にすることができる。

【0075】

<第2の強化学習>

次に、第1の強化学習に続いて、バックラッシュ加速補正部110におけるバックラッシュ加速補正パラメータを対象として強化学習(第2の強化学習)を行う場合について説明する。

前述したように、第1の強化学習により算出された、最適なバックラッシュ補正パラメータをバックラッシュ補正部109に設定したうえで、第2の強化学習を行う。

【0076】

バックラッシュ加速補正パラメータを対象とした第2の強化学習に際しては、例えば、台形型加速に基づく補正を適用することができる。台形型加速に基づく補正とは、各軸の速度指令の符号が反転したとき、又はその直前から、図6に示すように、台形の形状に基づいて「しばらくの時間」補正を行う方式を意味する。

図6に示す台形型加速に基づく補正を適用する場合、加速量 V_0 、計算開始時間TZ、加速時間T0、加速時間T1、及び減衰係数 α は、それぞれ速度指令の符号反転直後の加速度 x の平方根 $\sqrt{|x|}$ に応じて、以下の式(1)~式(6)のように定められる。ここで、 a_i 、 b_i 、 c_i ($i=1, 2, 3$)は、バックラッシュ加速補正のためのパラメータである。なお、加速時間T2については、加速度 x の平方根 $\sqrt{|x|}$ に依存しない定数パラメータである。

$$V_0 = a_1 \cdot \sqrt{|x|} \cdot H_{1/2}(b_1 - \sqrt{|x|}) + a_1 \cdot b_1 \cdot H_{1/2}(\sqrt{|x|} - b_1) + c_1 \quad (式1)$$

$$TZ = a_2 \cdot \sqrt{|x|} \cdot H_{1/2}(b_2 - \sqrt{|x|}) + a_2 \cdot b_2 \cdot H_{1/2}(\sqrt{|x|} - b_2) + c_2 \quad (式2)$$

$$T_0 = a_3 \cdot \sqrt{|x|} \cdot H_{1/2}(b_3 - \sqrt{|x|}) + a_3 \cdot b_3 \cdot H_{1/2}(\sqrt{|x|} - b_3) + c_3 \quad (式3)$$

$$\begin{aligned}
 T1 &= a_4 \cdot \sqrt{|x|} \cdot H_{1/2}(b_4 - \sqrt{|x|}) \\
 &+ a_4 \cdot b_4 \cdot H_{1/2}(\sqrt{|x|} - b_4) \\
 &+ c_4
 \end{aligned}
 \tag{式4}$$

$$\begin{aligned}
 \alpha &= a_5 \cdot \sqrt{|x|} \cdot H_{1/2}(b_5 - \sqrt{|x|}) \\
 &+ a_5 \cdot b_5 \cdot H_{1/2}(\sqrt{|x|} - b_5) \\
 &+ c_5
 \end{aligned}
 \tag{式5}$$

ここで、 $H_{1/2}(x)$ は、以下の値をとるヘヴィサイドの階段関数である。

$$\begin{aligned}
 H_{1/2}(x) &= 0 & (x < 0) \\
 H_{1/2}(0) &= 1/2 \\
 H_{1/2}(x) &= 1 & (x > 0)
 \end{aligned}$$

10

(式6)

【0077】

具体的には、加速量 V_0 を例にとると、図7Aのグラフに示すように、加速量 V_0 は、加速度 x の平方根 $\sqrt{|x|}$ の関数となっている。また、加速量 V_0 は、 $0 < \sqrt{|x|} < b_1$ においては、傾きが a_1 で y 切片の値が c_1 の一次関数となっており、 $\sqrt{|x|} > b_1$ においては、加速量 V_0 は、定数 $a_1 \cdot b_1 + c_1$ となっている。

【0078】

なお、例えば、加速量 V_0 と加速度 x の平方根 $\sqrt{|x|}$ との関係は、上記以外に、図7B又は図7Cのグラフに示す関係でも良い。

a_i 、 b_i ($6 \leq i \leq 15$)、 c_i ($6 \leq i \leq 21$) が、バックラッシュ加速補正のためのパラメータであって、例えば、加速量 V_0 と加速度 x の平方根 $\sqrt{|x|}$ との関係が図7Bのグラフに示す関係となる場合、 $0 < \sqrt{|x|} < b_6$ の範囲において、グラフの傾きを a_6 (> 0)、 y 切片を c_6 (> 0) とすると

20

$$V_0 = a_6 \sqrt{|x|} + c_6$$

となる。

$b_6 < \sqrt{|x|} < b_7$ の範囲において、グラフの傾きが a_7 (> 0)、 y 切片が c_7 (> 0) であって、 $a_7 \neq a_6$ 、 $c_7 \neq c_6$ とすると、

$$V_0 = a_7 \sqrt{|x|} + c_7$$

となる。

なお、 $\sqrt{|x|} = b_6$ のとき、 $a_6 \times b_6 + c_6 = a_7 \times b_6 + c_7 = c_8$ となる。

30

$b_7 < \sqrt{|x|}$ の範囲においては、 $c_9 > 0$ とすると、

$$V_0 = c_9$$

と常に定数となる。

これを、上記と同様に、ヘヴィサイドの階段関数を用いて表すと、

$$\begin{aligned}
 V_0 &= (a_6 \sqrt{|x|} + c_6) \cdot H_{1/2}(b_6 - \sqrt{|x|}) \\
 &+ (a_7 \sqrt{|x|} + c_7) \cdot H_{1/2}(\sqrt{|x|} - b_6) \cdot H_{1/2}(b_7 - \sqrt{|x|}) \\
 &+ c_9 \cdot H_{1/2}(\sqrt{|x|} - b_7)
 \end{aligned}
 \tag{式7}$$

となる。

【0079】

V_0 に限らず、 TZ 、 $T0$ 、 $T1$ 、 α についても、以下のように、これと同様の関数で表わすことが可能である。すなわち、

40

$$\begin{aligned}
 TZ &= (a_8 \sqrt{|x|} + c_{10}) \cdot H_{1/2}(b_8 - \sqrt{|x|}) \\
 &+ (a_9 \sqrt{|x|} + c_{11}) \cdot H_{1/2}(\sqrt{|x|} - b_8) \cdot H_{1/2}(b_9 - \sqrt{|x|}) \\
 &+ c_{12} \cdot H_{1/2}(\sqrt{|x|} - b_9)
 \end{aligned}
 \tag{式8}$$

$$\begin{aligned}
 T0 &= (a_{10} \sqrt{|x|} + c_{13}) \cdot H_{1/2}(b_{10} - \sqrt{|x|}) \\
 &+ (a_{11} \sqrt{|x|} + c_{14}) \cdot H_{1/2}(\sqrt{|x|} - b_{10}) \cdot H_{1/2}(b_{11} - \sqrt{|x|}) \\
 &+ c_{15} \cdot H_{1/2}(\sqrt{|x|} - b_{11})
 \end{aligned}
 \tag{式9}$$

$$T1 = (a_{12} \sqrt{|x|} + c_{16}) \cdot H_{1/2}(b_{12} - \sqrt{|x|})$$

50

$$\begin{aligned}
 & + (a_{13} \sqrt{|x|} + c_{17}) \cdot H_{1/2}((\sqrt{|x|} - b_{12})(b_{13} - \sqrt{|x|})) \\
 & + c_{18} \cdot H_{1/2}(\sqrt{|x|} - b_{13})
 \end{aligned}
 \tag{式 1}$$

0)

$$\begin{aligned}
 \alpha = & (a_{14} \sqrt{|x|} + c_{19}) \cdot H_{1/2}(b_{14} - \sqrt{|x|}) \\
 & + (a_{15} \sqrt{|x|} + c_{20}) \cdot H_{1/2}((\sqrt{|x|} - b_{14})(b_{15} - \sqrt{|x|})) \\
 & + c_{21} \cdot H_{1/2}(\sqrt{|x|} - b_{15})
 \end{aligned}
 \tag{式 1}$$

1)

【0080】

10

また、 a_i 、 b_i (16 i 25)、 c_i (22 i 37) が、バックラッシュ加速補正のためのパラメータであって、加速量 V_0 と加速度 x の平方根 $\sqrt{|x|}$ との関係が図 7 C のグラフに示す関係となる場合、 $0 < \sqrt{|x|} < b_{16}$ の範囲において、グラフの傾きを a_{16} (> 0)、 y 切片を c_{22} (> 0) とすると、

$$V_0 = a_{16} \sqrt{|x|} + c_{22}$$

となる。

$b_{16} < \sqrt{|x|} < b_{17}$ の範囲において、グラフの傾きが a_{17} (< 0)、 y 切片が c_{23} (> 0) であって、 $c_{23} \neq c_{22}$ とすると、

$$V_0 = a_{17} \sqrt{|x|} + c_{23}$$

なお、 $\sqrt{|x|} = b_{16}$ のとき、 $a_{16} \times b_{16} + c_{22} = a_{17} \times b_{16} + c_{23} = c_{25}$ となる。

$b_{17} < \sqrt{|x|}$ の範囲においては、 $c_{24} > 0$ とすると、

$$V_0 = c_{24}$$

と常に定数となる。

これを、上記と同様に、ヘヴィサイドの階段関数を用いて表すと、

$$\begin{aligned}
 V_0 = & (a_{16} \sqrt{|x|} + c_{22}) \cdot H_{1/2}(b_{16} - \sqrt{|x|}) \\
 & + (a_{17} \sqrt{|x|} + c_{23}) \cdot H_{1/2}((\sqrt{|x|} - b_{16})(b_{17} - \sqrt{|x|})) \\
 & + c_{24} \cdot H_{1/2}(\sqrt{|x|} - b_{17})
 \end{aligned}
 \tag{式 12}$$

12)

30

となる。

【0081】

V_0 に限らず、 TZ 、 T_0 、 T_1 、 α についても、以下のように、これと同様の関数で表わすことが可能である。すなわち、

$$\begin{aligned}
 TZ = & (a_{18} \sqrt{|x|} + c_{26}) \cdot H_{1/2}(b_{18} - \sqrt{|x|}) \\
 & + (a_{19} \sqrt{|x|} + c_{27}) \cdot H_{1/2}((\sqrt{|x|} - b_{18})(b_{19} - \sqrt{|x|})) \\
 & + c_{28} \cdot H_{1/2}(\sqrt{|x|} - b_{19})
 \end{aligned}
 \tag{式 13}$$

13)

$$\begin{aligned}
 T_0 = & (a_{20} \sqrt{|x|} + c_{29}) \cdot H_{1/2}(b_{20} - \sqrt{|x|}) \\
 & + (a_{21} \sqrt{|x|} + c_{30}) \cdot H_{1/2}((\sqrt{|x|} - b_{20})(b_{21} - \sqrt{|x|})) \\
 & + c_{31} \cdot H_{1/2}(\sqrt{|x|} - b_{21})
 \end{aligned}
 \tag{式 14}$$

14)

$$\begin{aligned}
 T_1 = & (a_{22} \sqrt{|x|} + c_{32}) \cdot H_{1/2}(b_{22} - \sqrt{|x|}) \\
 & + (a_{23} \sqrt{|x|} + c_{33}) \cdot H_{1/2}((\sqrt{|x|} - b_{22})(b_{23} - \sqrt{|x|})) \\
 & + c_{34} \cdot H_{1/2}(\sqrt{|x|} - b_{23})
 \end{aligned}
 \tag{式 15}$$

15)

$$\alpha = (a_{24} \sqrt{|x|} + c_{35}) \cdot H_{1/2}(b_{24} - \sqrt{|x|})$$

50

$$\begin{aligned}
 & + (a_{25} \sqrt{|x|} + c_{36}) \cdot H_{1/2}((\sqrt{|x|} - b_{24})(b_{25} - \sqrt{|x|})) \\
 & + c_{37} \cdot H_{1/2}(\sqrt{|x|} - b_{25}) \quad (\text{式 16})
 \end{aligned}$$

と表わすことが可能である。

【0082】

このように、バックラッシュ加速補正部 110 は、速度指令の符号反転直後の加速度 x の平方根 $\sqrt{|x|}$ に基づいて、(式 1) から (式 16) によりそれぞれ求められる、加速量 V_0 、計算開始時間 T_Z 、加速時間 T_0 、加速時間 T_1 、及び減衰係数 α 、並びに加速時間 T_2 に基づいて、図 6 に示す台形の形状に基づいて「しばらくの時間」補正を行うように動作する。

10

【0083】

第 2 の強化学習において、機械学習装置 200 は、サーボ制御装置 100 におけるバックラッシュ加速補正部 110 がバックラッシュ加速補正量を作成する際に用いるバックラッシュ加速補正パラメータ並びに学習時の位置指令に応じた制御を実行することで取得されるサーボ制御装置 100 の位置偏差情報を含むサーボ状態を状態 s として、当該状態 s に係る、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 の調整を行動 a として選択する価値 Q を学習する。

【0084】

バックラッシュ加速補正パラメータを対象とした強化学習においては、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 の調整（修正）を、行動 a として選択する価値 Q を学習する。

20

【0085】

機械学習装置 200 は、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 及び加速時間 T_2 に基づいて、学習時の加工プログラムをサーボ制御装置 100 に実行させた場合の位置指令と位置フィードバック値から得られる位置偏差値の集合と、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 と、を含む状態情報 s を観測して、次の行動 a （バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 の調整）を決定する。

機械学習装置 200 は、行動 a をするたびに報酬が返ってくる。機械学習装置 200 は、例えば、将来にわたっての報酬の合計が最大になるように最適な行動 a を試行錯誤的に探索する。そうすることで、機械学習装置 200 は、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 に基づいて、学習時の加工プログラムをサーボ制御装置 100 に実行させることで得られるサーボ制御装置 100 の位置偏差情報を含むサーボ状態を含む状態情報 s に対して、最適な行動 a 、すなわち、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 を選択することが可能となる。

30

【0086】

これにより、第 2 の強化学習において、機械学習装置 200 により学習された価値関数 Q に基づいて、或る状態 s に係るバックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 及び加速時間 T_2 に対して適用される行動 a のうち、 Q の値が最大となるような行動 a を選択することで、位置偏差が最小になるような行動 a （すなわち、バックラッシュ加速補正部 110 の最適なバックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間 T_2 を選択すること）が可能となる。

40

【0087】

なお、第 2 の強化学習における、状態情報取得部 201、学習部 202（報酬出力部 2021、価値関数更新部 2022、及び行動情報生成部 2023）、行動情報出力部 203、価値関数記憶部 204、及び最適化行動情報出力部 205 の処理は、前述した第 1 の強化学習時と同等の処理を行う。すなわち、機械学習装置 200 は、第 1 の強化学習についての前述の説明において、バックラッシュ補正部 109 をバックラッシュ加速補正部 110 に読み替え、バックラッシュ補正パラメータ P_0 をバックラッシュ加速補正パラメータ a_i 、

50

b_i 、 c_i 、及び加速時間 T_2 に読み替えた強化学習を行う。

以上のように、本発明に係る機械学習装置 200 を利用することで、サーボ制御装置 100 のバックラッシ加速補正部 110 のバックラッシ加速補正パラメータ調整を簡易化することができる。

【0088】

以上、サーボ制御装置 100、機械学習装置 200 に含まれる機能ブロックについて説明した。

これらの機能ブロックを実現するために、サーボ制御装置 100、及び機械学習装置 200 のそれぞれは、CPU (Central Processing Unit) 等の演算処理装置を備える。また、サーボ制御装置 100、及び機械学習装置 200 のそれぞれは、アプリケーションソフトウェアや OS (Operating System) 等の各種の制御用プログラムを格納した HDD (Hard Disk Drive) 等の補助記憶装置や、演算処理装置がプログラムを実行する上で一時的に必要とされるデータを格納するための RAM (Random Access Memory) といった主記憶装置も備える。

【0089】

そして、サーボ制御装置 100、及び機械学習装置 200 のそれぞれにおいて、演算処理装置が補助記憶装置からアプリケーションソフトウェアや OS を読み込み、読み込んだアプリケーションソフトウェアや OS を主記憶装置に展開させながら、これらのアプリケーションソフトウェアや OS に基づいた演算処理を行なう。また、この演算結果に基づいて、各装置が備える各種のハードウェアを制御する。これにより、本実施形態の機能ブロックは実現される。つまり、本実施形態は、ハードウェアとソフトウェアが協働することにより実現することができる。

【0090】

機械学習装置 200 については機械学習に伴う演算量が多いため、例えば、パーソナルコンピュータに GPU (Graphics Processing Units) を搭載し、GP-GPU (General-Purpose computing on Graphics Processing Units) と呼ばれる技術により、GPU を機械学習に伴う演算処理に利用するようにすると高速処理できるようになるのでよい。更には、より高速な処理を行うために、このような GPU を搭載したコンピュータを複数台用いてコンピュータ・クラスターを構築し、このコンピュータ・クラスターに含まれる複数のコンピュータにて並列処理を行うようにしてもよい。

【0091】

上記のサーボ制御装置のサーボ制御部及び機械学習装置に含まれる各構成部は、ハードウェア、ソフトウェア又はこれらの組み合わせにより実現することができる。また、上記のサーボ制御装置に含まれる各構成部のそれぞれの協働により行なわれるサーボ制御方法も、ハードウェア、ソフトウェア又はこれらの組み合わせにより実現することができる。ここで、ソフトウェアによって実現されるとは、コンピュータがプログラムを読み込んで実行することにより実現されることを意味する。

【0092】

プログラムは、様々なタイプの非一時的なコンピュータ可読媒体 (non-transitory computer readable medium) を用いて格納され、コンピュータに供給することができる。非一時的なコンピュータ可読媒体は、様々なタイプの実体のある記録媒体 (tangible storage medium) を含む。非一時的なコンピュータ可読媒体の例は、磁気記録媒体 (例えば、フレキシブルディスク、磁気テープ、ハードディスクドライブ)、光磁気記録媒体 (例えば、光磁気ディスク)、CD-ROM (Read Only Memory)、CD-R、CD-R/W、半導体メモリ (例えば、マスク ROM、PROM (Programmable ROM)、EPROM (Erasable PROM)、フラッシュ ROM、RAM (random access memory)) を含む。また、プログラムは、様々なタイプの一時的なコンピュータ可読媒体 (transitory computer readable medium) によってコンピュータに供給されてもよい。一時的なコンピュータ可読媒体の例は、電気信号、光信号、及び電磁波を含む。一時的なコンピュータ可読媒体は、電線及び光ファイバ等の有線通信路、又は無線通信路を介して、プログラムをコンピュータに供給できる

。

【0093】

<本実施形態の動作>

次に、図8のブロック図、図9のフローチャートを参照してバックラッシュ補正パラメータを対象とする第1の強化学習時の機械学習装置200の動作について説明をする。

【0094】

バックラッシュ補正パラメータを対象とする第1の強化学習時には、図8の太線で示されるように、機械学習装置200は、バックラッシュ補正パラメータ P_0 の調整情報を含む行動情報をバックラッシュ補正部109に対して出力し、当該行動情報に基づいて、学習時の加工プログラムをサーボ制御装置100に実行させて、位置偏差情報を観測する。

10

【0095】

第1の強化学習は、図9に示すように、ステップS11において、学習時の加工プログラムを実行することで、状態情報取得部201がサーボ制御装置100から状態情報 s を取得する。取得した状態情報は、価値関数更新部202や行動情報生成部2023に対して出力される。上述したように、この状態情報 s は、Q学習における状態に相当する情報であり、ステップS11時点での、バックラッシュ補正部109のバックラッシュ補正パラメータ P_0 が含まれる。こうして、バックラッシュ補正パラメータ P_0 が初期値であるときの、学習時の加工プログラムにより指定される所定の送り速度且つ円の加工形状に対応する位置偏差の集合 $PD(s)$ が取得される。

【0096】

20

最初にQ学習を開始する時点での減算器102から状態 s_0 における位置偏差の値 $PD(s_0)$ は、学習時の加工プログラムでサーボ制御装置100を動作させることで得られる。なお、学習時の加工プログラムにおいては、サーボモータ302の回転速度は前述したとおり、送り速度を変更し複数の送り速度が用意されている。位置指令作成部101は、学習時の加工プログラムにより指定された所定の加工形状で、例えば、円、四角、角R付き四角(a square with quarter arc)の加工形状で、送り速度を変えて位置指令を順次出力する。例えば、所定の送り速度で円の加工形状に対応する位置指令値が位置指令作成部101から出力され、減算器102は位置指令値と積分器108から出力される検出位置との差を位置偏差 $PD(s_0)$ として機械学習装置200に出力する。

【0097】

30

ステップS12において、行動情報生成部2023は新たな行動情報 a を生成し、生成した新たな行動情報 a を、行動情報出力部203を介してサーボ制御装置100に対して出力する。行動情報生成部2023は前述した方策に基づいて、新たな行動情報 a を出力する。なお、行動情報 a を受信したサーボ制御装置100は、受信した行動情報に基づいて現在の状態 s に係るバックラッシュ補正部109のバックラッシュ補正パラメータ P_0 を修正した状態 s' により、サーボモータを含む工作機械を駆動させる。

【0098】

ステップS13において、状態情報取得部201は、減算器102から新たな状態 s' における位置偏差 $PD(s')$ を取得する。こうして、状態情報取得部201は、バックラッシュ補正部109から状態 s' に係るバックラッシュ補正パラメータ P_0 に基づいて、学習時の加工プログラムを前記サーボ制御装置に実行させた場合における、位置偏差の集合 $PD(s')$ を取得する。取得した状態情報は、報酬出力部2021に対して出力される。

40

【0099】

ステップS14において、報酬出力部2021は、状態 s' における位置偏差の値 $f(PD(s'))$ と状態 s における位置偏差の値 $f(PD(s))$ との大小関係を判断し、 $f(PD(s')) > f(PD(s))$ の場合には、ステップS15において、報酬を負の値とする。 $f(PD(s')) < f(PD(s))$ の場合には、ステップS16において、報酬を正の値とする。 $f(PD(s')) = f(PD(s))$ の場合には、ステップS17において、報酬をゼロとする。なお、報酬の負の値、正の値について重みづけを行

50

うようにしてもよい。

【0100】

ステップS15、ステップS16及びステップS17の何れかが終了すると、ステップS18において、この何れかのステップにて算出された報酬の値に基づいて、価値関数更新部2022が、価値関数記憶部204に記憶している行動価値関数Qを更新する。そして、再度ステップS11に戻り、上述した処理を繰り返すことにより、行動価値関数Qは適切な値に収束していく。なお、上述した処理を、所定回数繰り返したことや、所定時間繰り返したことを条件として処理を終了するようにしてもよい。

なお、ステップS18はオンライン更新を例示しているが、オンライン更新に替えてバッチ更新又はミニバッチ更新に置き換えてもよい。

10

【0101】

以上、図8及び図9を参照して説明した第1の強化学習により、本実施形態では、機械学習装置200を利用することで、バックラッシュ補正パラメータ P_0 の最適な値を容易に取得することができる。

【0102】

次に、図10のフローチャートを参照して、最適化行動情報出力部205による最適化行動情報の生成時の動作について説明をする。

まず、ステップS21において、最適化行動情報出力部205は、価値関数記憶部204に記憶している行動価値関数Qを取得する。行動価値関数Qは、上述したように価値関数更新部2022がQ学習を行うことにより更新したものである。

20

【0103】

ステップS22において、最適化行動情報出力部205は、この行動価値関数Qに基づいて、最適化行動情報を生成し、生成した最適化行動情報をサーボ制御装置100のバックラッシュ補正部109に対して出力する。

このようにして、第1の強化学習を終了する。

【0104】

次にバックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 及び加速時間T2を対象とする第2の強化学習時の機械学習装置200の動作について説明をする。

機械学習装置200は、バックラッシュ加速補正パラメータを対象とした第2の強化学習時にも、図9のフローチャートと同様の動作フローにより、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及びT2の最適な値を容易に取得することができる。

30

【0105】

なお、バックラッシュ加速補正パラメータを対象とした第2の強化学習時には、図11の太線で示されるように、バックラッシュ補正部109は、加算器103に対して、第1の強化学習で最適化されたバックラッシュ補正パラメータ P_0 を適用して、バックラッシュ補正量を出力する。

そして、機械学習装置200は、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 及び加速時間T2の調整情報を含む行動情報をバックラッシュ加速補正部110に対して出力し、当該行動情報に基づいて、学習時の加工プログラムをサーボ制御装置100に実行させて、位置偏差情報を観測する。

40

【0106】

なお、第2の強化学習における、状態情報取得部201、学習部202（報酬出力部2021、価値関数更新部2022、及び行動情報生成部2023）、行動情報出力部203、価値関数記憶部204、及び最適化行動情報出力部205の動作は、前述した第1の強化学習時と同等の動作を行う。すなわち、機械学習装置200は、第1の強化学習についての前述の動作説明において、バックラッシュ補正部109をバックラッシュ加速補正部110に読み替え、バックラッシュ補正パラメータ P_0 をバックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及び加速時間T2に読み替えた強化学習の動作を行う。

このようにして、第2の強化学習時に、バックラッシュ加速補正パラメータ a_i 、 b_i 、 c_i 、及びT2の最適な値を容易に取得することができる。

50

バックラッシ加速補正パラメータを対象とした第２の強化学習後には、図１２の太線で示されるように、バックラッシ加速補正部１１０は、加算器１０６に対して、第２の強化学習で最適化されたバックラッシ加速補正パラメータ a_i 、 b_i 、 c_i 、及び T_2 を適用して、バックラッシ加速補正量を出力する。

【０１０７】

上述した実施形態は、本発明の好適な実施形態ではあるが、上記実施形態のみに本発明の範囲を限定するものではなく、本発明の要旨を逸脱しない範囲において種々の変更を施した形態での実施が可能である。

【０１０８】

< サーボ制御装置が機械学習装置を備える変形例 >

10

上述した実施形態では、機械学習装置２００を、サーボ制御装置１００とは別体の装置により構成したが、機械学習装置２００の機能の一部又は全部をサーボ制御装置１００により実現するようにしてもよい。

【０１０９】

< システム構成の自由度 >

上述した実施形態では、機械学習装置２００とサーボ制御装置１００とが１対１の組として通信可能に接続されているが、例えば１台の機械学習装置２００が複数のサーボ制御装置１００とネットワーク４００を介して通信可能に接続され、各サーボ制御装置１００の機械学習を実施するようにしてもよい。

その際、機械学習装置２００の各機能を、適宜複数のサーバに分散する、分散処理システムとしてもよい。また、クラウド上で仮想サーバ機能等を利用して、機械学習装置２００の各機能を実現してもよい。

20

また、複数の同じ型名、同一仕様、又は同一シリーズのサーボ制御装置１００－１～１００－ n とそれぞれ対応する複数の機械学習装置２００－１～２００－ n があった場合に、各機械学習装置２００－１～２００－ n における学習結果を共有するように構成するようにしてもよい。そうすることで、より最適なモデルを構築することが可能となる。

【符号の説明】

【０１１０】

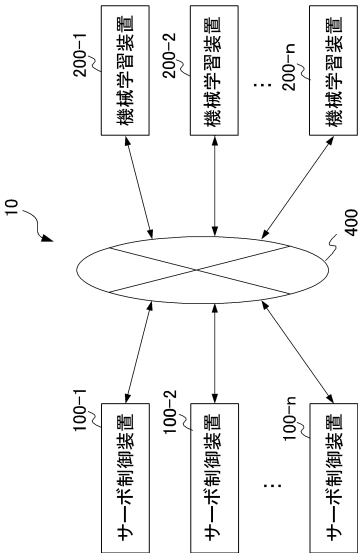
- １０ サーボ制御システム
- １００ サーボ制御装置
- １０１ 位置指令作成部
- １０２ 減算器
- １０３ 加算器
- １０４ 位置制御部
- １０５ 減算器
- １０６ 加算器
- １０７ 速度制御部
- １０８ 積分器
- １０９ バックラッシ補正部
- １１０ バックラッシ加速補正部
- ２００ 機械学習装置
- ２０１ 状態情報取得部
- ２０２ 学習部
- ２０３ 行動情報出力部
- ２０４ 価値関数記憶部
- ２０５ 最適化行動情報出力部
- ３００ 制御対象
- ４００ ネットワーク
- ２０２１ 報酬出力部
- ２０２２ 価値関数更新部

30

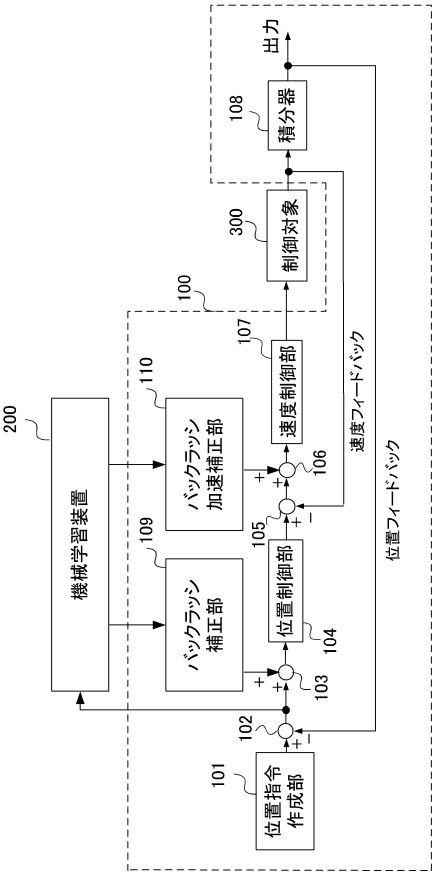
40

50

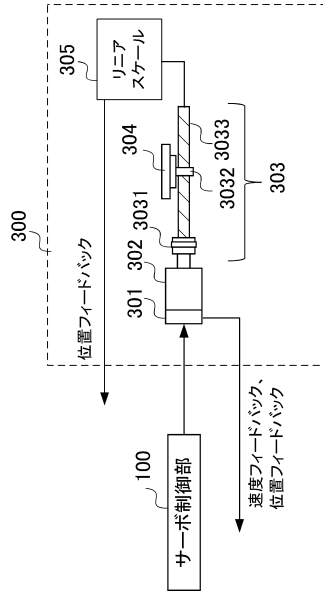
【 図 1 】



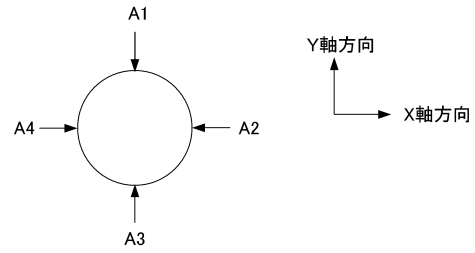
【 図 2 】



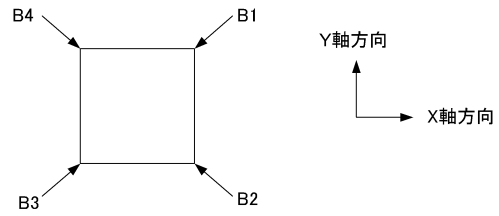
【図 3】



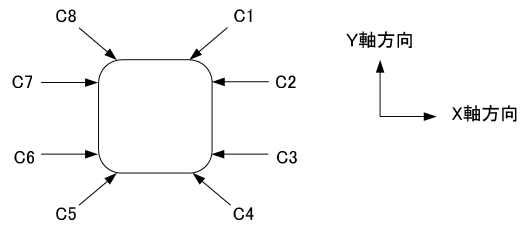
【図 4 A】



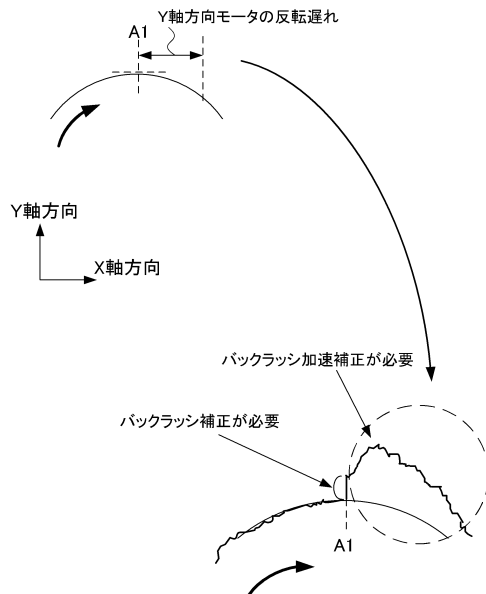
【図 4 B】



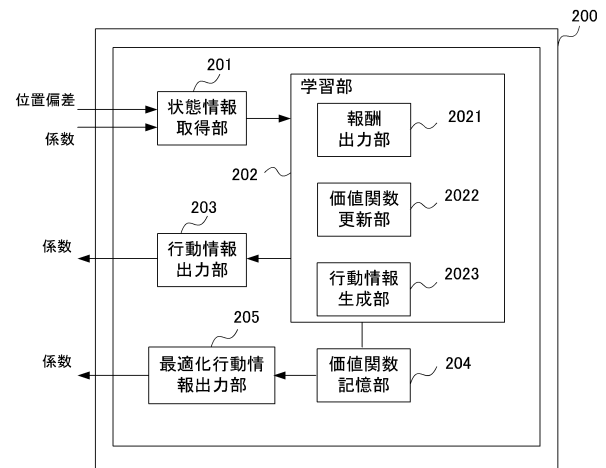
【図 4 C】



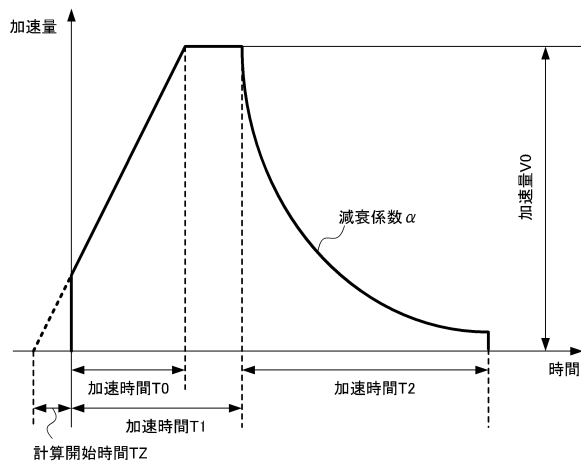
【図 4 D】



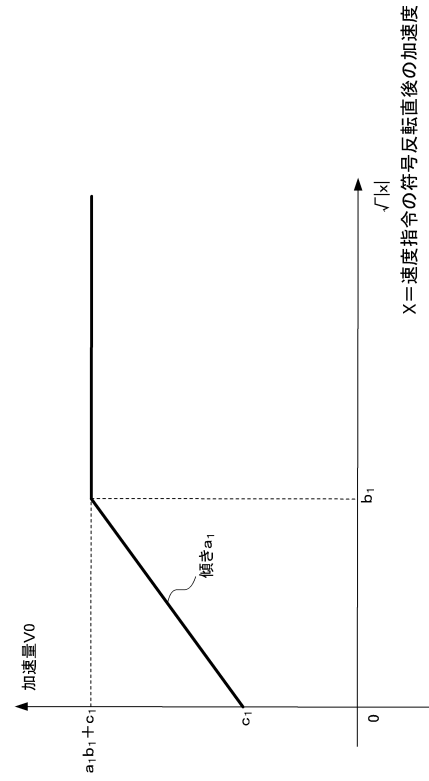
【図 5】



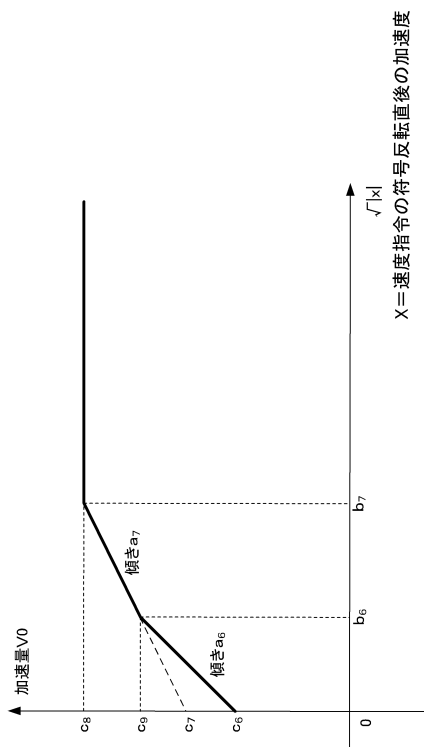
【図 6】



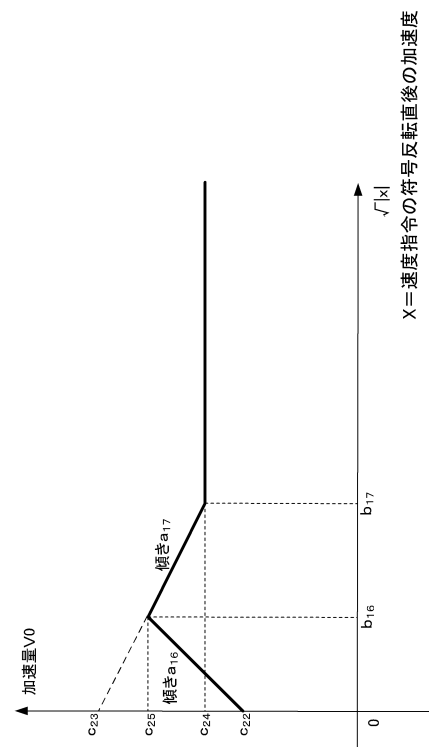
【図 7 A】



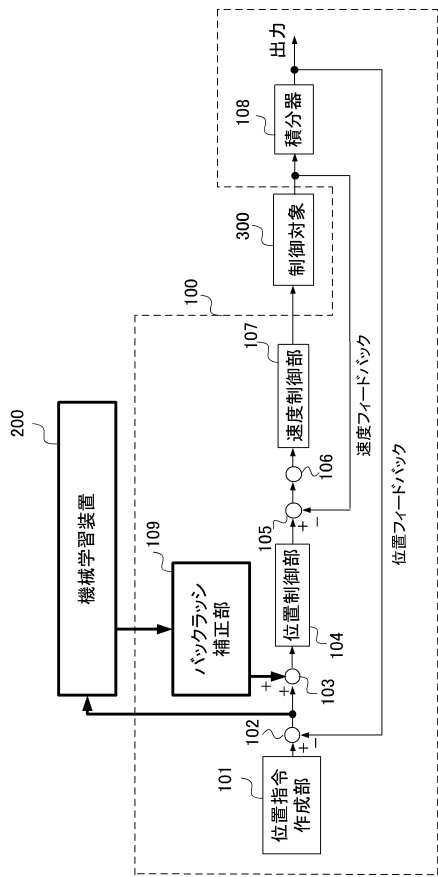
【図 7 B】



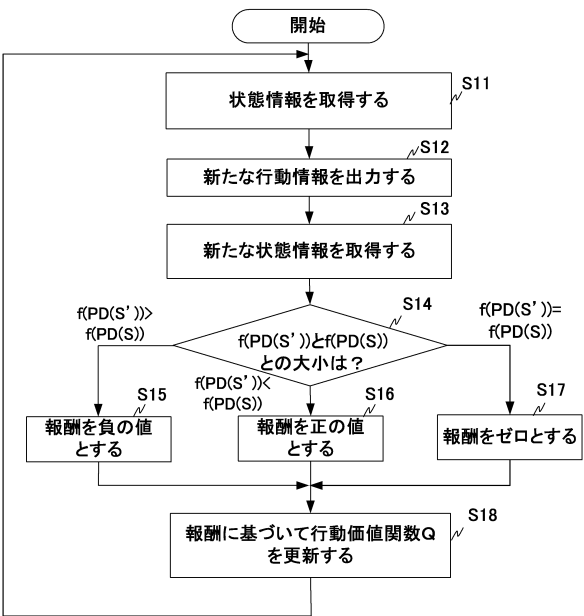
【図 7 C】



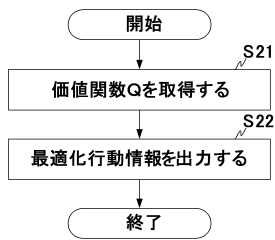
【図 8】



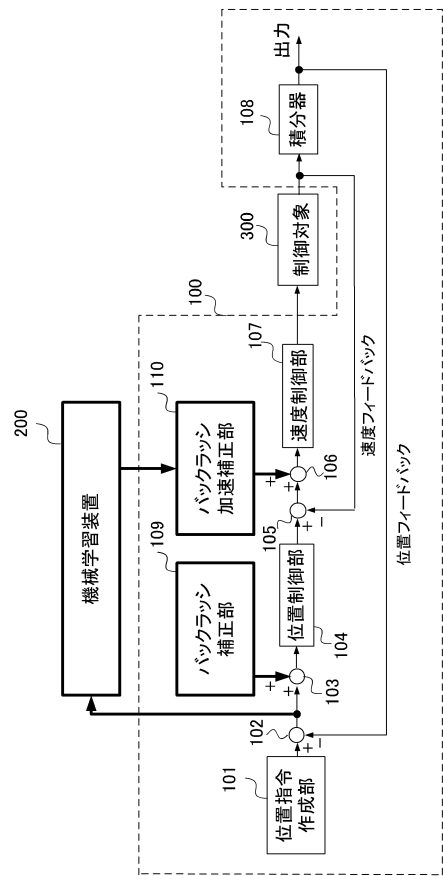
【図 9】



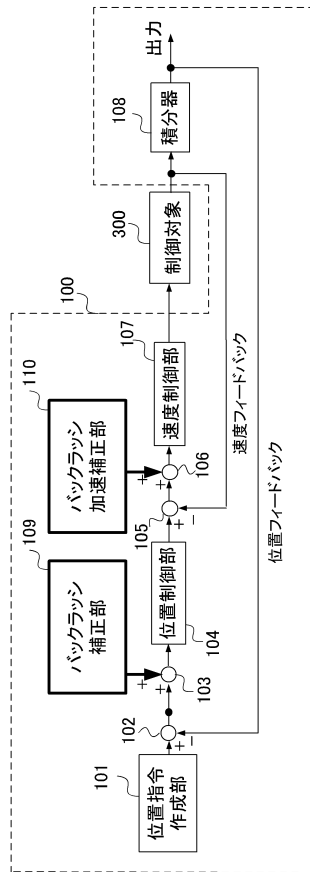
【図 10】



【図 11】



【図 12】



フロントページの続き

(72)発明者 中邨 勉

山梨県南都留郡忍野村忍草字古馬場3580番地 ファナック株式会社内

(72)発明者 相澤 智之

山梨県南都留郡忍野村忍草字古馬場3580番地 ファナック株式会社内

審査官 影山 直洋

(56)参考文献 特開2012-093982(JP,A)

特開2017-064837(JP,A)

(58)調査した分野(Int.Cl., DB名)

G05B 19/404

G05B 13/02