

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年3月19日(19.03.2020)



(10) 国際公開番号

WO 2020/054345 A1

(51) 国際特許分類:
G06N 3/08 (2006.01)

(21) 国際出願番号: PCT/JP2019/032837

(22) 国際出願日: 2019年8月22日(22.08.2019)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2018-168864 2018年9月10日(10.09.2018) JP

(71) 出願人: 日立オートモティブシステムズ株式会社(HITACHI AUTOMOTIVE SYSTEMS, LTD.) [JP/JP]; 〒3128503 茨城県ひたちなか市高場 2 5 2 0 番地 Ibaraki (JP).

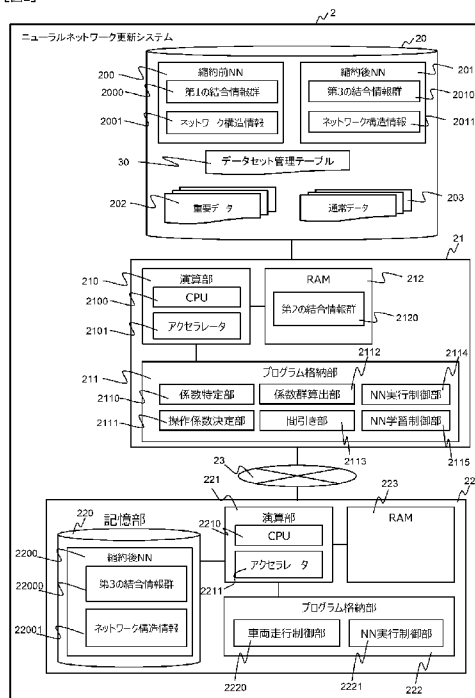
(72) 発明者: 木谷 光博 (KITANI, Mitsuhiro); 〒1008280 東京都千代田区丸の内一丁目 6 番 6 号 株式会社日立製作所内 Tokyo (JP). 北原 倫理 (KITAHARA, Michimasa); 〒1008280 東京都千代田区丸の内一丁目 6 番 6 号 株式会社日立製作所内 Tokyo (JP). 祖父江 恒夫(SOBUE, Tsuneco); 〒1008280 東京都千代田区丸の内一丁目 6 番 6 号 株式会社日立製作所内 Tokyo (JP).

(74) 代理人: 特許業務法人サンネクスト国際特許事務所(SUNNEXT INTERNATIONAL PATENT OFFICE); 〒1400002 東京都品川区東品川二丁目 3 番 1 2 号 シーフオートスクエア センタービルディング 1 6 階 Tokyo (JP).

(54) Title: ELECTRONIC CONTROL DEVICE AND NEURAL NETWORK UPDATE SYSTEM

(54) 発明の名称: 電子制御装置、ニューラルネットワーク更新システム

[図2]



2... NEURAL NETWORK UPDATE SYSTEM
30... DATA SET MANAGEMENT TABLE
200... PRE-CONTRACTION NN
201, 2200... POST-CONTRACTION NN
202... IMPORTANT DATA
203... ORDINARY DATA
210, 221... COMPUTATION UNIT
211, 222... PROGRAM STORAGE UNIT
220... STORAGE UNIT
2000... FIRST JOINT INFORMATION GROUP
2001, 2011, 22001... NETWORK STRUCTURE INFORMATION
2010, 22000... THIRD JOINT INFORMATION GROUP
2101, 2211... ACCELERATOR
2110... COEFFICIENT IDENTIFICATION UNIT
2111... OPERATIONAL COEFFICIENT DETERMINATION UNIT
2112... COEFFICIENT GROUP CALCULATION UNIT
2113... CULLING UNIT
2114, 2221... NN EXECUTION CONTROL UNIT
2115... NN LEARNING CONTROL UNIT
2120... SECOND JOINT INFORMATION GROUP
2220... VEHICLE TRAVEL CONTROL UNIT

(57) Abstract: This electronic control device stores a neural network containing a plurality of neurons and joint information group including a plurality of items of joint information tying the neurons together, wherein the neural network includes a third joint information group that is created by a culling unit deleting at least one item of joint information from a first joint information group for determining an output value when an input value has been given, on the basis of the first joint information group and a second joint information group that is a plurality of items of joint information a degree

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

添付公開書類:

一 国際調査報告(条約第21条(3))

of impact of which on predetermined data given as the input value exceeds a predetermined value.

(57) 要約: 電子制御装置は、複数のニューロンと、ニューロン同士を紐付ける結合情報を複数含む結合情報群とを含むニューラルネットワークを格納する電子制御装置であって、ニューラルネットワークは、入力値が与えられたときに出力値を決定するための第1の結合情報群と、入力値として与えられた所定のデータに対する影響度が所定値を上回る複数の結合情報である第2の結合情報群とに基づき、間引き部が第1の結合情報群から少なくとも1つの結合情報を削除することで作成される第3の結合情報群を含むニューラルネットワークである。

明 細 書

発明の名称：

電子制御装置、ニューラルネットワーク更新システム

技術分野

[0001] 本発明は、電子制御装置、およびニューラルネットワーク更新システムに関する。

背景技術

[0002] 近年、自動運転システムの開発が活発化している。自動運転システムにおいて、複雑な走行環境を走行するためには、カメラ、レーザレーダ、ミリ波レーダ等の様々なセンサからの情報を基にした自車両周辺環境をセンシングする「認識」や、センサで検出された自車両周辺のオブジェクトが今後どのように行動するのかを推定する「認知」や、認識と認知の結果に基づき今後の自車両の行動を計画する「判断」といった機能の高度化が必要になる。そのため、これらの機能に対して、ニューラルネットワークやディープラーニングなどのAI (Artificial Intelligence) モデルを導入することにより更なる高度化が期待されている。車両用の電子制御装置 (ECU : Electronic Control Unit) において、車両走行制御で用いることが可能な処理周期でリアルタイムにニューラルネットワークの演算処理を実行する場合、CPUだけではなくFPGA (Field-Programmable Gate Array)、ASIC (Application Specific Integrated Circuit)、GPU (Graphics Processing Unit) などのハードウェアデバイスを追加で用いることが多く、ECUの消費電力の増大が課題となる。そのため、ニューラルネットワークの演算量を減らして、ニューラルネットワーク実行時の消費電力を低減する必要がある。特許文献1には、入力した画像のニューロンデータとパラメータとに基づいてそれぞれ算出した畳み込み量を各第1のメモリ領域にそれぞれ保持する処理と、前記各第1のメモリ領域に保持した前記畳み込み量に対して間引き処理をそれぞれ行った間引き後の畳み込み量を各第2のメモリ領域にそれ

ぞれ保持する処理とを第1の複数層において行う第1の認識処理と、前記第2のメモリ領域に保持した前記間引き後の畳み込み量の全てに対して第3のメモリ領域に保持した重みをそれぞれ積算した出力結果を各第4のメモリ領域にそれぞれ保持する処理を第2の複数層において行う第2の認識処理とを制御する認識制御部と、前記第1の複数層と前記第2の複数層とに含まれる各層について、各ニューロンデータのサイズであるニューロンデータサイズと、各パラメータのサイズであるパラメータサイズとをそれぞれ計算するメモリ量計算部と、前記第4のメモリ領域に保持した前記出力結果に基づき計算した前記出力結果の誤差の勾配を第5のメモリ領域に保持するとともに、前記メモリ量計算部が計算した前記第2の複数層に含まれる各層のニューロンデータサイズとパラメータサイズと大小関係に基づき、前記第5のメモリ領域に保持した前記出力結果の誤差の勾配又は前記第2の複数層における前層の第6のメモリ領域に保持した誤差の勾配に基づいて計算した前記第2の複数層における次層への誤差の勾配を各第6のメモリ領域に保持した後、前記第2の複数層における次層へのパラメータの誤差の勾配を前記各第3のメモリ領域にそれぞれ保持する前記第2の複数層における第1の学習処理と、前記メモリ量計算部が計算した前記第1の複数層に含まれる各層のニューロンデータサイズとパラメータサイズとの大小関係に基づき、前記第2の複数層の最終層の第6のメモリ領域又は前記第1の複数層における前層の第7のメモリ領域に保持した誤差の勾配に基づいて計算した次層へのパラメータの誤差の勾配を各第7のメモリ領域に保持した後、前記第1の複数層における次層への誤差の勾配を前記各第2のメモリ領域にそれぞれ保持する前記第1の複数層間における第2の学習処理とを制御する学習制御部と、を有することを特徴とする画像認識装置が開示されている。

先行技術文献

特許文献

[0003] 特許文献1：日本国特開2018-18350号公報

発明の概要

発明が解決しようとする課題

- [0004] 特許文献 1 に記載されている発明では、重み係数を削除した場合に高精度な推論が要求されるデータに対する推論精度が、重み係数削除前と比べて低下する場合がある。

課題を解決するための手段

- [0005] 本発明の第 1 の態様による電子制御装置は、複数のニューロンと、前記ニューロン同士を紐付ける結合情報を複数含む結合情報群とを含むニューラルネットワークを格納する電子制御装置であって、前記ニューラルネットワークは、入力値が与えられたときに出力値を決定するための第 1 の前記結合情報群と、前記入力値として与えられた所定のデータに対する影響度が所定値を上回る複数の前記結合情報である第 2 の結合情報群とに基づき、間引き部が前記第 1 の結合情報群から少なくとも 1 つの結合情報を削除することで作成される第 3 の結合情報群を含むニューラルネットワークである。

本発明の第 2 の態様によるニューラルネットワーク更新システムは、サーバ、および電子制御装置を備え、複数のニューロンと、前記ニューロン同士を紐付ける結合情報を複数含む結合情報群とを含むニューラルネットワークを更新するニューラルネットワーク更新システムであって、前記サーバは、記憶部と、演算部と、を備え、前記記憶部は、入力値が与えられたときに出力値を決定するための第 1 の前記結合情報群を格納し、前記演算部は、所定のデータが前記入力値として与えられた場合に、前記所定のデータに対する影響度が所定値を上回る複数の前記結合情報である第 2 の結合情報群を特定する係数特定部と、前記第 1 の結合情報群と、前記第 2 の結合情報とに基づき、前記第 1 の結合情報群から少なくとも 1 つの結合情報を削除して第 3 の結合情報群を作成する間引き部と、を備え、前記電子制御装置は、前記第 3 の結合情報群を含むニューラルネットワークを格納する。

発明の効果

- [0006] 本発明によれば、高精度な推論が要求されるデータに対する推論精度の低下を限定的に留めてニューラルネットワークの演算量を低減することができ

る。

図面の簡単な説明

[0007] [図1] A I モデルの構造の一例を示す図

[図2] 第 1 の実施の形態におけるニューラルネットワーク更新システム 2 の構成図

[図3] データセット管理テーブル 3 0 の一例を示す図

[図4] サーバ 2 1 によるニューラルネットワークの縮約処理を示すフローチャート

[図5] 図 4 における S 4 0 2 の詳細を示すフローチャート

[図6] 変形例 1 における、データセット管理テーブル 3 1 の一例を示す図

[図7] 変形例 4 における、図 3 に示すステップ S 4 0 2 の詳細を示す図

[図8] 変形例 4 における、ステップ S 6 0 2 の処理の一例を示す図

[図9] 変形例 5 における、図 3 に示すステップ S 4 0 2 の詳細を示す図

[図10] 図 9 に示す処理の一例を示す図

[図11] 変形例 6 における、図 3 に示すステップ S 4 0 2 の詳細を示す図

[図12] ステップ S 9 0 2 の処理の一例を示す図

[図13] 変形例 6 におけるシステム構成図

[図14] 第 2 の実施の形態におけるニューラルネットワーク更新システム 2 の構成図

[図15] 縮約率補正テーブル 2 0 4 の一例を示す図

発明を実施するための形態

[0008] 以下、本発明の実施の形態について図面を参照して説明する。なお、以下に示す各実施の形態では、同じ構成要素や同じ処理内容等については、同一の番号を記載してその説明を簡略にする。各実施の形態では、人工知能処理を用いた A I モデルを備えた電子制御装置、サーバ、全体システムについての説明において、A I モデルの一例として、ニューラルネットワークを用いて説明する。ただし A I モデルは、機械学習、ディープラーニング、および強化学習に関するモデルであってもよい。すなわち、近似や間引き処理によ

り、最適化できる様々なAIモデルに対して、以下に説明する各実施の形態を適用できる。なお以下に説明する実施の形態では、モデルの最適化は、重み係数やニューロンを間引く方法を例に説明し、その処理のことを「ニューラルネットワークの縮約」として記載する。また以下では、「ニューラルネットワーク」を「NN」と記載することもある。

[0009] 一第1の実施の形態一

以下、図1～図5を参照して、ニューラルネットワーク更新システムの第1の実施の形態を説明する。第1の実施の形態では、高精度な推論が要求されるデータ（以下、「重要データ」とも呼ぶ）に深く関連するニューラルネットワークの結合情報を特定し、その情報を用いて結合情報の間引きを行う。なお、間引き処理は結合情報そのものを削除する方法や、値を0に変更する方法が考えられる。

[0010] 図1は、AIモデルの構造の一例を示す図である。図1に示すように、ニューラルネットワークモデル1は、入力層10と、中間層11と、出力層12を含み、各層はそれぞれI個、J個、K個の演算ユニット100を持つ。それぞれの演算ユニット100は、図1の符号101や符号102に示すように、演算ユニット100間の結合情報を基に接続される。入力層10に情報が入力されると、結合情報が加味されて中間層11内で伝搬されて、最終的に出力層12より予測結果に相当する情報が出力される。演算ユニット100間の接続に関する結合情報を本実施の形態では「ネットワーク構造情報」と呼ぶ。演算ユニット100は前述の「ニューロン」に相当する。

[0011] 結合情報は結合係数とバイアスから構成され、その結合係数を係数とする演算が行われつつ情報は入力層10から出力層12に伝搬される。AIモデルの演算で使用される結合係数のことを本実施の形態では「重み係数」と呼ぶ。また本実施の形態では、説明を簡略するためにバイアスを除外して結合情報が重み係数のみから構成されるときとして説明する。すなわち本実施の形態では、結合情報と重み係数は同一である。ただし結合情報にバイアスが含まれる構成も本発明に含まれる。

- [0012] また本実施の形態では、個別の重み係数を識別するために、それぞれの重み係数に便宜的にIDを付与する。また重み係数の値を「重み係数値」や「重み係数の値」と呼ぶ。すなわち以下では、「IDが11の重み係数の重み係数値は××」のように表現する。演算の内容は、結合情報に含まれる層の種類によって特定される。結合情報に含まれる層は、たとえば、畳込み層、バッチ正規化、活性化関数、プーリング層、全結合層、LSTM (Long Short Term Memory) 層などである。
- [0013] なお演算ユニット100の数や、中間層11を構成する層の数は、実施の形態とは無関係であるため、任意の値とする。また、AIモデルの構造も特段の限定はされず、たとえば演算ユニット100間の結合に再帰性や双方向性を持つものであってもよい。さらに教師有り、教師無しのいずれの機械学習モデル、および強化学習モデル等、いずれのAIモデルに対しても、本実施の形態を適用可能である。これは、高精度な推論が要求される優先度の高いデータに対する推論精度に影響を与えることなく、AIモデルの重み係数や演算ユニットを間引いてAIモデルを縮約できるためである。ただし重み係数の重み係数値をゼロに変更することも、重み係数を間引くことに含まれる。
- [0014] 図2は、第1の実施の形態におけるニューラルネットワーク更新システム2の構成図である。ニューラルネットワーク更新システム2は、記憶部20と、サーバ21と、電子制御装置22とから構成される。図2に示すように、サーバ21と電子制御装置22は、ネットワーク23を介して接続される。なお図2では記憶部20はサーバ21と直接に接続されているが、記憶部20がサーバ21とネットワーク23を介して接続されてもよい。
- [0015] 本実施の形態では、サーバ21にてニューラルネットワークの縮約を行い、電子制御装置22はその縮約されたニューラルネットワークをサーバからダウンロードして更新する。これにより、自動運転などのアプリケーションにおいて、ニューラルネットワークを用いた演算処理を実行する。ただし、本実施の形態はこのような構成に限定されない。すなわちサーバ21を介さ

ず、電子制御装置 22 を備える車両に閉じてニューラルネットワークの追加学習と更新を行うアーキテクチャになる場合など、上述の機能の全部または一部を電子制御装置 22 に搭載する構成としてもよい。

[0016] 記憶部 20 は、不揮発性の記憶領域、たとえばフラッシュメモリである。記憶部 20 には、ニューラルネットワークの学習、生成、および縮約処理において、サーバ 21 が使用するニューラルネットワーク情報やデータセットが格納される。具体的には記憶部 20 には、縮約前ニューラルネットワーク 200 と、縮約後ニューラルネットワーク 201 と、重要データ 202 と、通常データ 203 とが含まれる。

[0017] 縮約前ニューラルネットワーク 200 は、縮約を適用する前のニューラルネットワークの情報を指し、第 1 の結合情報群 2000 とネットワーク構造情報 2001 とから構成される。第 1 の結合情報群 2000 は縮約適用前のニューラルネットワークの重み係数の重み係数値である。ネットワーク構造情報 2001 は縮約適用前のニューラルネットワークのネットワーク構造情報である。縮約前 NN 200 はたとえば、図 1 に示すニューラルネットワークモデル 1 の全体である。第 1 の結合情報群 2000 はたとえば、図 1 に示した NN モデル 1 から抽出した全ての重み係数の重み係数値である。またネットワーク構造情報 2001 はたとえば、NN モデル 1 から全ての重み係数の重み係数値を除いた情報である。

[0018] 縮約後ニューラルネットワーク 201 は、縮約を適用した後のニューラルネットワークの情報であり、第 3 の結合情報群 2010 とネットワーク構造情報 2011 とから構成される。第 3 の結合情報群 2010 は縮約適用後のニューラルネットワークの重み係数の重み係数値を指す。ネットワーク構造情報 2011 は縮約適用前のニューラルネットワークのネットワーク構造情報を指す。なお、縮約適用前後でニューラルネットワークのネットワーク構造情報に相違がない場合は、ネットワーク構造情報 2001 とネットワーク構造情報 2011 は同一のものとなる。

[0019] 重要データ 202 および通常データ 203 は、あらかじめ作成されたデー

タ群であり、ニューラルネットワークの構築や推論のテストに用いられる。重要データ 202 および通常データ 203 に含まれるそれぞれデータは、たとえば走行中の車両が収集したセンサ出力と、前方を走行する車両との相対距離と、前方を走行する車両が同一レーンを走行しているか否かの情報の組み合わせである。

[0020] 重要データ 202 は、高精度な推論が要求されるデータであり、たとえば車両走行制御に直接関連し、推論の誤りを可能な限り低減したいデータ群である。重要データ 202 はたとえば、相対距離が近く、かつ同一レーンを走行している場合のデータである。通常データ 203 は、重要データ 202 以外の、車両走行制御に直接関係しないデータ群である。通常データ 203 はたとえば、相対距離が非常に遠いデータや、異なるレーンを走行している場合のデータである。あるデータが重要データ 202 および通常データ 203 のどちらに分類されるかは、次に説明するデータセット管理テーブル 30 を参照してサーバ 21 が演算処理により決定してもよいし、サーバ 21 のオペレータがあらかじめ定められたルールに従い決定してもよい。

[0021] 図 3 はデータセット管理テーブル 30 の一例を示す図である。データセット管理テーブル 30 は複数のレコードから構成され、各レコードは、データ種別名 300、優先度判定条件 301、優先度 302、操作係数 303 から構成される。データ種別名 300 はデータを識別するために便宜的に設けた名称である。したがってデータ種別名 300 は識別可能であればどのような文字列でもよい。優先度判定条件 301 は、優先度を判定するための条件であり、換言すると当該レコードへの該当性を示す条件である。

[0022] 優先度 302 は数値が大きいほど優先度が高い、すなわちより高精度な推論が要求されることを示す。それぞれのデータは、この優先度の値により重要データ 202 または通常データ 203 に分類されてもよい。操作係数 303 は、縮約時に用いられる情報であり、たとえば間引きの対象となる重み係数、すなわち間引きの対象となる重み係数の ID を決定するために使用される。

- [0023] 優先度判定条件 301 は、相対距離 3010 と先行車フラグ 3011 とから構成される。相対距離 3010 は自車とオブジェクトとの相対距離を表す情報であり、相対距離の値が小さいほど、優先度 302 の値は大きい、すなわち本データの優先度は高いと判定する傾向にある。先行車フラグ 3011 は、対象のオブジェクトが自車から見て同一走行レーン上の先行車にあたるか否かを表すフラグ情報である。
- [0024] 先行車フラグ 3011 が TRUE の場合は、優先度の値は大きくなる傾向にある。図 3 の 1 行目に示す例では、先行車フラグ 3011 がオンかつ相対距離 3010 が 30メートル以内の場合は、オブジェクト検出用データ 1 と判断され、優先度 302 は 3、操作係数は 2 に設定される。また図 3 の最終行に示す例では、先行車フラグ 3011 が FALSE の場合は相対距離 3010 が 500メートル以内は全てオブジェクト検出用データ 4 と判断され、優先度 302 は 1、操作係数は 1 に設定される。
- [0025] 図 2 の説明に戻る。サーバ 21 は演算部 210 と、プログラム格納部 211 と、RAM 212 とを備える。演算部 210 は、中央演算装置である CPU (Central Processing Unit) 2100 をハードウェアとして少なくとも備えており、CPU 2100 はプログラム格納部 211 に格納されたプログラムに従ってニューラルネットワークの学習、生成、および縮約処理を実行する。ただしプログラム格納部 211 に格納されたプログラムと同様の機能をハードウェア回路として実現してもよい。さらに、プログラム格納部 211 に格納される全てのプログラムの機能がハードウェア回路として実現される場合は、サーバ 21 はプログラム格納部 211 を備えなくてもよい。
- [0026] ニューラルネットワークの学習・生成・縮約を高速化するため、演算部 210 にアクセラレータ 2101 が含まれてもよい。アクセラレータ 2101 は、FPGA、ASIC、および GPU 等のハードウェアデバイスが考えられる。
- [0027] プログラム格納部 211 は、フラッシュメモリ等の不揮発性半導体メモリを備える。プログラム格納部 211 には、演算係数特定部 2110 と、操作

係数決定部 2 1 1 1 と、係数群算出部 2 1 1 2 と、間引き部 2 1 1 3 と、N N 実行制御部 2 1 1 4 と、N N 学習制御部 2 1 1 5 とを実現するためのプログラムが格納される。これらのプログラムは演算部 2 1 0 により実行される。これらのプログラムの動作は図 4 ～図 5 を参照して後に説明する。またプログラムの構成は任意であり、どのようにプログラムの結合や分離がされてもよい。たとえば、1 つのプログラムが演算係数特定部 2 1 1 0、操作係数決定部 2 1 1 1、係数群算出部 2 1 1 2、間引き部 2 1 1 3、N N 実行制御部 2 1 1 4、および N N 学習制御部 2 1 1 5 に相当する機能の全てを兼ね備えてもよい。

[0028] R A M 2 1 2 は揮発性半導体メモリである。R A M 2 1 2 には、演算部 2 1 0 がプログラムを実行するためのデータ、たとえば第 2 の結合情報群 2 1 2 0 が格納される。第 2 の結合情報群 2 1 2 0 は、ニューラルネットワークを縮約する際の判定条件として使用される。第 2 の結合情報群 2 1 2 0 は、「重要関連 W H I」とも呼ばれる。詳しくは後述する。

[0029] ネットワーク 2 3 は、電子制御装置 2 2 を備える車両がサーバ 2 1 へアクセスするための通信網を表す。アクセス手段はたとえばセルラーや無線 L A N などの方式が考えられる。これら無線手段以外に O B D (On-Board Diagnostics) や O B D 2 (On-Board Diagnostics2) を介して有線手段で接続してもよい。

[0030] 電子制御装置 2 2 は、記憶部 2 2 0 と、演算部 2 2 1 と、プログラム格納部 2 2 2 と、R A M 2 2 3 とを備える。記憶部 2 2 0 は、フラッシュメモリなどの不揮発性の記憶装置である。記憶部 2 2 0 には縮約後ニューラルネットワーク 2 2 0 0 が格納される。縮約後 N N 2 2 0 0 は、第 3 の結合情報群 2 2 0 0 0 とネットワーク構造情報 2 2 0 0 1 とから構成される。電子制御装置 2 2 に格納される縮約後 N N 2 2 0 0 は、記憶部 2 0 に格納される縮約後 N N 2 0 1 と名称は同一であるが、データそのものは必ずしも同一ではない。ただし出荷直後の初期状態では、縮約後 N N 2 2 0 0 と縮約後 N N 2 0 1 とは同一である。

[0031] 演算部221は、中央演算処理装置であるCPU2210を備える。CPU2210はプログラム格納部222に格納されたプログラムに従ってニューラルネットワークを用いた推論処理やニューラルネットワークの学習処理を実行する。演算部221は、これら推論や学習処理を高速化するため、アクセラレータ2101を備えてもよい。アクセラレータ2101はFPGA、ASIC、GPU等のハードウェアデバイスが考えられる。プログラム格納部222は、フラッシュメモリ等の不揮発性半導体メモリを備える。プログラム格納部222には、車両走行制御部2220と、NN実行制御部2221とを実現するためのプログラムが格納される。

[0032] 車両走行制御部2220は、認識、認知、および判断などの車両の走行を制御するために必要な機能を有する。なお、車両走行制御部2220が単独で車両の走行を制御するための認識、認知、および判断を実行しなくてもよく、たとえば車両に搭載された他の機器やネットワーク23を介して接続される他の機器と分担・連携してもよい。

[0033] NN実行制御部2221は、記憶部220に格納された縮約後NN2200を用いて、推論処理を実行する。図示はしていないが、電子制御装置22は推論処理だけではなく、縮約後ニューラルネットワーク2200の学習機能をさらに有してもよい。電子制御装置22が学習機能を有することにより、サーバ21と連携することなく、電子制御装置22が単独でニューラルネットワークの追加学習と更新が可能になる。

[0034] (動作)

図4～図5を参照して、サーバ21によるニューラルネットワークの縮約を説明する。図4はサーバ21によるニューラルネットワークの縮約処理を示すフローチャートである。まずステップS400において、NN学習制御部2115は、学習済みの重み係数値WLEを備えるニューラルネットワーク、すなわち縮約前のニューラルネットワークを生成する。

[0035] そして、ステップS401において、記憶部20に格納されているデータをデータセット管理テーブル30を参照して重要データ202を選定する。

この場合に、たとえば優先度が3以上のデータが重要データ202に分類され、優先度が2以下のデータが通常データ203に分類される。なおこの分類はデータの選定はサーバ21のオペレータが行ってもよいし、サーバ21が演算により選別してもよい。また、データがすでに分類済みの場合はS401の処理を省略してもよい。

[0036] 続いてステップS402において、演算係数特定部2110は、学習済み重み係数値WLEのなかで、重要データ202に関与する重み係数を特定し、そのIDの集合を重要関連WHIとして記録する。すなわち重要関連WHIとは、重要データ202に関与する重み係数のID番号が列挙されたものである。S402の詳細は次の図5で説明する。

[0037] 続くステップS403において、操作係数決定部2111は、データセット管理テーブル30と重要データ202に基づいて操作係数WCOEFを算出する。操作係数WCOEFはそれぞれの重み係数に対応して存在し、全ての操作係数WCOEFの初期値は同一、たとえばゼロである。図3に示す例において最上段のオブジェクト検出用データ1に分類される重要データ202を用いて特定された重み係数に対応する操作係数WCOEFには、そのデータの操作係数303の値である「2」が加算される。多くの重要データ202により特定される重み係数に対応する操作係数WCOEFほど大きな値を有することになる。

[0038] ステップS404において、係数群算出部2112は、重要関連WHIと操作係数WCOEFを用いて、ニューラルネットワークにおける重み係数の間引き適用箇所の判定に用いる評価係数WNEWを算出する。評価係数WNEWも、それぞれの重み係数に対して存在する。評価係数WNEWは、それぞれの重み係数に対して、換言するとそれぞれの重み係数のIDに対して、たとえば次の式1のように算出される。

$$WNEW = D(WHI) * Val(ID) * WCOEF \quad \dots$$

(式1)

[0039] ただしD(WHI)は、処理対象の重み係数のIDが重要関連WHIに含

まれる場合は「1」を返し、処理対象の重み係数のIDが重要関連WHIに含まれない場合はゼロを返す関数である。またVal(ID)は、処理対象のIDである重み係数の重み係数値を返す関数である。詳しくは後述するが、評価係数NEWは間引きの判定処理で参照される値であり、値が大きいほど間引きがされにくくなる。

[0040] 続くステップS405において、間引き部2113は評価係数NEWの値に基づいて間引きの対象となる重み係数を決定する。そしてステップS406において、学習済み重み係数値WLEから間引き対象の重み係数の重み係数値をゼロにして、新たなニューラルネットワークを生成する。ただし、重み係数の行成分または列成分のすべてがゼロになった場合は、その行または列を削除してもよい。

[0041] さらに、各演算ユニット100（ニューロン）に接続される全ての重み係数のセット単位で削除しても良い。これは、演算ユニット100に接続される各重み係数の二乗和をそれぞれ算出して、その値が小さいものから順に重み係数のセットを削除する方法が考えられる。生成されたニューラルネットワークは縮約後NN201として記憶部20に格納され、電子制御装置22に送信されて記憶部220に保存される。なお、間引き適用箇所は1つ以上とする。

[0042] その後ステップS407において、ステップS406にて新たに生成された新たなニューラルネットワークにおける重み係数を再学習、すなわちファインチューニングによって最適化する。このファインチューニングにより推論精度の向上を図ることが可能となる。ただしステップS407は必須ではなくS406の実行が完了したら図4に示す処理を終了してもよい。

[0043] 図5は、図4におけるS402の詳細を示すフローチャートである。図5に示す処理の実行主体は係数特定部2110である。係数特定部2110はまずS500において、NN実行制御部2114に、重要データ202をステップS400において生成したニューラルネットワークに入力して推論を実行させ、ニューラルネットワークの各演算ユニットの出力OHNUを算出

する。続くステップS501では係数特定部2110は、出力OHNUに基づき、活性化度合いが高い演算ユニットNH1を特定する。本処理で特定した演算ユニットは重要データ202に大きく関与している演算ユニットであると判断する。

[0044] ここで活性化度合いが高いか否かは、出力OHNUが閾値ACTTH以上か否かによって判定する方法が考えられる。別の方法としてはニューラルネットワークの各層毎の全ての出力OHNUのうち、値が大きいものから順に上位から一定の割合を活性化度合いが高いと判定する方法が考えられる。

[0045] その後、ステップS502において、係数特定部2110は、ステップS501で特定した演算ユニットNH1に対して、演算ユニットNH1と接続される重み係数を特定し、特定した重み係数を重要関連WH1とすることによって、本処理フローを終了する。ステップS502の処理の一例を示すと、演算ユニットNH1が図1に示す演算ユニット100の場合に、以下のよう判断する。すなわち、演算ユニット100に接続される重み係数は結合情報101と結合情報102である。このように、それぞれの演算ユニットに接続される重み係数は、ニューラルネットワークのネットワーク構造情報から一意に特定される。

[0046] 上述した第1の実施の形態によれば、次の作用効果が得られる。

(1) ニューラルネットワーク更新システム2は、サーバ21および電子制御装置22を備える。ニューラルネットワーク更新システム2は、複数のニューロン100と、ニューロン同士を紐付ける図1の符号101や符号102に示す結合情報を複数含む結合情報群とを含むニューラルネットワークを更新する。サーバ21は、記憶部20と、演算部210と、を備える。記憶部20は、入力値が与えられたときに出力値を決定するための第1の結合情報群2000を格納する。演算部210は、所定のデータが入力値として与えられた場合に、所定のデータに対する影響度が所定値を上回る複数の結合情報である第2の結合情報群2120を特定する係数特定部2110と、第1の結合情報群2000と、第2の結合情報群2120とに基づき、第1の

結合情報群 2000 から少なくとも 1 つの結合情報を削除して第 3 の結合情報群 2010 を作成する間引き部 2113 と、を備える。電子制御装置 22 は、第 3 の結合情報群 2010 を含む縮約後 NN 2200 を格納する。

[0047] 第 3 の結合情報群 2010 は第 1 の結合情報群 2000 に比べて結合情報が削除されているので演算量、すなわち消費電力を低減できる。さらに、高精度な推論が要求される重要データに対する影響度を考慮して、削除する結合情報を決定するので、重要データに対する推論精度の低下を限定的に留めることができる。

[0048] (2) 所定のデータとは、車両走行に関する優先度が所定値を上回るデータである。そのため車両の走行に関して高精度な推論が要求されるデータに対する推論制度の低下を限定的に留めることができる。

[0049] (3) 間引き部 2113 は、第 2 の結合情報群 2120 に含まれるそれぞれの結合情報に対応する操作係数の集合である操作係数群を決定し、それぞれの操作係数を第 1 の結合情報群 2000 に含まれる対応する結合情報に評価係数を乗算して評価係数群とし、評価係数群のうち所定値を上回る評価係数について、対応する結合情報を第 1 の結合情報群 2000 の中から特定し、当該特定された結合情報以外の結合情報を第 1 の結合情報群 2000 から削除して、第 3 の結合情報群 2010 を作成する。このように、高精度な推論が要求される重要データに対する影響度が所定値を上回る結合情報は削除しないので、重要データに対する推論精度の低下を限定的に留めることができる。

[0050] (4) 係数特定部 2110 は、複数のニューロン 100 のうち、所定のデータが与えられた場合の出力値が所定値を上回る場合に、当該出力値の決定に関与したニューロン 100 を特定し、当該特定されたニューロン 100 に紐付く結合情報を、第 2 の結合情報群 2120 を構成する結合情報として特定する。

[0051] (変形例 1)

記憶部 20 に格納されるデータ、すなわち重要データ 202 および通常デ

ータ203は、前方を走行する車両との相対距離が含まれていなくてもよい。記憶部20に格納されるデータには、高精度な推論が要求されるか否かを判断可能な情報が含まれていればよい。たとえば記憶部20に格納されるデータに物体の種別と高さが含まれる場合は、データセット管理テーブル30を図6に示すデータセット管理テーブル31に変更する。

[0052] 図6に示すデータセット管理テーブル31は、第1の実施の形態におけるデータセット管理テーブル30と比べて優先度判定条件301の内容が異なり、優先度判定条件301は識別クラス3012とオブジェクト高さ3013とから構成される。識別クラス3012は物体の種別を示し、図3に示す例では物体が車両、歩行者、および自転車のいずれかの場合は優先度を高く設定される。オブジェクト高さ3013は物体の高さを示し、高いほうが優先度が高く設定される。

[0053] なお、記憶部20に格納されるデータに様々な情報が含まれる場合は、このデータを使用するアプリケーションに応じて優先度判定条件301が変更されてもよい。すなわちデータごとに優先度302および操作係数303を決定できればよく、優先度判定条件301は他の情報を用いてもかまわない。たとえば優先度判定条件301が、自己位置推定、センサフュージョン、マップフュージョン、フリースペース検知、オブジェクトの移動予測、および走行軌道計画などAIが使用される他の自動運転ロジックに関する条件でもよい。

[0054] (変形例2)

電子制御装置22の記憶部220に縮約前NN220がさらに格納され、縮約後NN2200に何らかの問題が発生した場合に縮約前NN220に切り替えて、ニューラルネットワークを用いた演算処理を実行してもよい。何らかの問題とはたとえば、縮約後NN2200の出力結果を用いたことで、衝突回避のため急ブレーキを動作せざるを得なくなることである。このような問題が生じた場合に縮約前NN220に切り替えることで、縮約処理を施したことによる車両走行への影響を排除することができる。

[0055] (変形例3)

評価係数 W_{NEW} の決定方法は、第1の実施の形態において示した式1を用いる方法に限定されない。重要データ202に關与する重み係数が間引きされにくくなるように、評価係数 W_{NEW} の値が大きくなるようにすればよい。たとえば、処理対象のIDである重み係数の重み係数値に対して操作係数 W_{COEF} などのような補正值を用いてビットシフト演算を行い評価係数 W_{NEW} を算出してもよい。また処理対象のIDである重み係数の重み係数値に操作係数 W_{COEF} などの補正值を加算して評価係数 W_{NEW} を算出してもよい。

[0056] (変形例4)

図7は変形例4における、図3に示すステップS402の詳細を示す図である。すなわち本変形例では、図5に示す処理の代わりに図6に示す処理が実行される。ただし第1の実施の形態と同一の処理には同一のステップ番号を付して説明を省略する。図7ではまずステップS500を実行し、第1の実施の形態と同様に出力 $OHNU$ を算出する。

[0057] 続くステップS601において、NN学習制御部2115は通常データ203をニューラルネットワークへの入力として推論を実行し、ニューラルネットワークの各演算ユニットの出力 $OLNU$ を算出する。そして、ステップS602において、演算係数特定部2110は、 $OHNU$ と $OLNU$ に基づき、たとえば $OHNU$ と $OLNU$ の差に基づき、活性化度合いが高い演算ユニット NHI を特定する。続くS502では、第1の実施の形態と同様に重要関連 WHI を特定して図7に示す処理を終了する。

[0058] 図8は、ステップS602の処理の一例を示す図である。図8(a)は、ステップS600における演算ユニットごとの $OHNU$ を示す図であり、図8(b)は、ステップS601における演算ユニットごとの $OLNU$ を示す図であり、図8(c)は演算ユニットごとに $OHNU$ と $OLNU$ の差分を示す図である。ただし図8(a)～(c)において横軸方向に異なる演算ユニットの出力が並んでおり、図8(a)～(c)の左右方向の位置が同一であ

れば同一の演算ユニットについて示している。

[0059] 図8では、それぞれの演算ユニットにIDを割り当てている。たとえばIDが「n11」の演算ユニットは図8(a)～図8(c)の左端に示されており、図8(a)と図8(b)の差が図8(c)である。そして図8(c)において、閾値WTHを設定し、閾値WTHよりも大きい演算ユニットを活性化度合いが高い演算ユニットNHIとして特定する。

[0060] (変形例5)

図9は変形例5における、図3に示すステップS402の詳細を示す図である。すなわち本変形例では、図5に示す処理の代わりに図9に示す処理が実行される。ただし第1の実施の形態と同一の処理には同一のステップ番号を付して説明を省略する。本変形例では、重要データ202に大きく関連する演算ユニットを特定せずに重み係数を特定する。

[0061] 図9ではまずステップS800において、NN学習制御部2115は、重要データ202を用いてニューラルネットワークの追加学習を行い、追加学習後の重み係数WHADDLEを算出する。ステップS800における追加学習で用いるデータは、縮約前NN200を生成する際に用いたデータとは異なるものである。

[0062] 続くステップS801において、係数特定部2110は、追加学習前の重み係数値WLEと追加学習後の重み係数値WHADDLEとに基づき、たとえば両者の重み係数値の差を演算して重要データ202に関与する重み関数を特定する。追加学習前の重み係数値WLEとは、図4のS400において算出した重み係数値WLEである。そして図9に示す処理を終了する。

[0063] 図10は、図9に示す処理の一例を示す図である。図10(a)は、縮約前NN200の重み係数のIDごとに重み係数値WLEを示す図であり、図10(b)は、ステップS800における重み係数のIDごとに重み係数値WHADDLEを示す図であり、図10(c)は重み係数ごとに重み係数値WHADDLEと重み係数値WLEの差分を示す図である。

[0064] たとえば、重み係数値WHADDLEと重み係数値WLEの差が所定の閾

値W T Hよりも大きい重み係数を重要データ202に大きく関連すると判断できる。またこの際に、重み係数値W H A D D L Eと重み係数値W L Eの差の絶対値を評価してもよい。さらに、両者の差が大きいものから順に上位から一定の割合を活性化度合いが高い重み係数と判定してもよい。

[0065] (変形例6)

図11は変形例6における、図3に示すステップS402の詳細を示す図である。すなわち本変形例では、図5に示す処理の代わりに図11に示す処理が実行される。ただし第1の実施の形態と同一の処理には同一のステップ番号を付して説明を省略する。本変形例では、変形例5と同様に重要データ202に大きく関連する演算ユニットを特定せずに重み係数を特定する。

[0066] 図11ではまずステップS900において、NN学習制御部2115は、重要データ202を用いてニューラルネットワークの学習を行い、重み係数値W H L Eを算出する。続くステップS901において、NN学習制御部2115は、通常データ203を用いてニューラルネットワークの学習を行い、重み係数値W L L Eを算出する。続くステップS902において、係数特定部2110は、学習済みの重み係数値W H L Eと重み係数値W L L Eに基づき、重要データ202に関与する重み係数を特定し、重要関連W H Iと名付ける。

[0067] なおステップS900とステップS901は、ステップS400におけるニューラルネットワークを生成する処理において実行されてもよい。また、ステップS400で使用したものと異なる重要データ202および通常データ203を用いてニューラルネットワークの追加学習を行い、これによりステップS900とステップS901の処理を実施してもよい。

[0068] ステップS902の処理の一例を図12に示す。図12(a)は、ステップS900における重み係数のIDごとの重み係数値W H L Eを示したものである。図12(b)は、ステップS901における重み係数のIDごとの重み係数値W H A D D L Eを示したものである。図12(c)は重み係数のIDごとの、図12(a)に示す重み係数値W H L Eと、図12(b)に示

す重み係数値W H A D D L Eとの差を示す図である。

[0069] 図12(c)において、IDごとの差分値が閾値W T H以上か否かの判定を行うことで各重み係数が重要データ202に大きく関連するか否かを判定する。ここで重み係数値W H L Eと重み係数値W L L Eの差分結果に対して絶対値を取るようにしてもよい。別の方法としては差分値が大きいものから順に上位から一定の割合を活性化度合いが高いと判定する方法が考えられる。

[0070] (変形例7)

図13は、変形例7におけるシステム構成を示す図である。本変形例では、車両で取得した走行データを活用してサーバ21と連携しながら電子制御装置22に対して縮約したニューラルネットワークを更新する。図13におけるニューラルネットワーク更新システム2は、図2で示したシステム構成と比べて、以下のように電子制御装置22の構成が異なる。

[0071] 本変形例では、記憶部220に重要データ2201がさらに保存される。重要データ2201は車両に搭載されるセンサが車両の周囲をセンシングして得られたデータのうち、後述の重要データ選定部2222によって抽出されたデータである。プログラム格納部222には重要データ選定部2222と、NN更新制御部2223と、重要データアップロード部2224とを実行するためのプログラムがさらに格納される。

[0072] 重要データ選定部2222は車両において外界センシング等によって収集したデータの中から車両走行制御に関連し優先度が高いと判定するデータを選定する。選定方法は図3で説明したものが考えられる。また別の方法として、事故発生には至らなかったが、衝突回避のため急ブレーキが動作した状況に関するセンシングデータを重要データとして選定することが考えられる。それ以外にもドライバによる手動運転時や、ニューラルネットワークを使用しない自動運転ロジックや運転支援ロジックを実行している際、ニューラルネットワークによる推論処理をバックグラウンドで実行しておき、手動運転時の出力結果やニューラルネットワークを用いないロジックの出力結果と

、ニューラルネットワークによる推論処理結果に乖離がある場合に、ニューラルネットワークへの入力データを重要データとして選定することが考えられる。

[0073] NN更新制御部2223は、サーバ21で新たに生成されたニューラルネットワークを取得し、記憶部220に格納される縮約後NN2200を更新する。重要データアップロード部2224は、重要データ選定部2222で選定されて、記憶部220にバッファリングされた重要データ2201を、ネットワーク23を介してサーバ21にアップロードする。

[0074] (変形例8)

上述した第1の実施の形態では、縮約処理において演算係数を削除する構成を説明した。しかしニューロン、すなわち演算ユニットを削除してもよい。演算ユニットを削除できるのは、その演算ユニットに接続される演算係数が全てゼロになった場合、またはその演算ユニットに接続される演算係数が全て削除された場合である。

[0075] 一第2の実施の形態一

図14～図15を参照して、ニューラルネットワーク更新システムの第2の実施の形態を説明する。以下の説明では、第1の実施の形態と同じ構成要素には同じ符号を付して相違点を主に説明する。特に説明しない点については、第1の実施の形態と同じである。本実施の形態では、主に、重み係数を間引く割合を決定する点で、第1の実施の形態と異なる。重み係数を間引く割合を本実施の形態では「縮約率」と呼ぶ。

[0076] 図14は、第2の実施の形態におけるニューラルネットワーク更新システム2の構成図である。図14に示す構成は、図2に示した構成と比べて、サーバ21のプログラム格納部211に縮約率補正部2116を実現するためのプログラムがさらに格納され、記憶部20に縮約率補正テーブル204がさらに格納される。縮約率補正部2116は、係数特定部2110によって特定された演算ユニットや重み係数情報と縮約率補正テーブル204とを用いて、間引き部2113で用いられる縮約率を補正する。縮約率補正テーブ

ル 204 は、縮約率の補正値を決定するために用いられる情報である。

[0077] 図 15 は、縮約率補正テーブル 204 の一例を示す図である。縮約率補正テーブル 204 は重要データに大きく関与する演算ユニット数 140 と縮約率 141 とを対応付けて管理するためのテーブルである。重要データ関連演算ユニット数 140 の欄には、重要データに大きく関与する演算ユニットの数が記述される。縮約率 141 の欄には、縮約時に使用される縮約率が記述される。縮約率とはニューラルネットワークの層単位で重み係数を間引く程度を表す割合情報である。たとえば縮約率の規定値が 20% に設定されている場合は、その層の重み係数を 20% 間引くことになる。

[0078] 縮約率 141 の例ではもともと設定されている縮約率に対して乗算や除算で補正することを想定して記述しているが、ビットシフト演算や加減算などの別の手段で設定するようにしても良い。また、縮約率を 10%、20% のように絶対値で設定するようにしてもよい。縮約率補正テーブル 204 を使用する場合、たとえば、ニューラルネットワークの任意の層において、係数特定部 2110 によって特定された演算ユニットの数 = 25 となった場合は、その層の縮約率は本テーブル情報を使って通常の縮約率から 2 で除算したものとなる。なお、本例では層単位での処理を記載しているが、層単位以外のニューラルネットワークにおける任意の処理単位であってもよい。

[0079] 上述した各実施の形態および変形例は、それぞれ組み合わせてもよい。上記では、種々の実施の形態および変形例を説明したが、本発明はこれらの内容に限定されるものではない。本発明の技術的思想の範囲内で考えられるその他の態様も本発明の範囲内に含まれる。

[0080] 次の優先権基礎出願の開示内容は引用文としてここに組み込まれる。

日本国特許出願 2018-168864 (2018 年 9 月 10 日出願)

符号の説明

[0081] 100 … 演算ユニット

2 … ニューラルネットワーク更新システム

20 … 記憶部

- 2 1 ... サーバ
- 2 2 ... 電子制御装置
- 2 3 ... ネットワーク
- 2 0 0 ... 縮約前ニューラルネットワーク
- 2 0 1 ... 縮約後ニューラルネットワーク
- 2 0 2 ... 重要データ
- 2 0 3 ... 通常データ
- 2 0 4 ... 縮約率補正テーブル
- 2 1 0 ... 演算部
- 2 2 0 ... 記憶部
- 2 2 1 ... 演算部
- 2 0 0 0 ... 第 1 の結合情報群
- 2 0 1 0 ... 第 3 の結合情報群
- 2 1 1 0 ... 係数特定部
- 2 1 1 1 ... 操作係数決定部
- 2 1 1 2 ... 係数群算出部
- 2 1 1 3 ... 間引き部
- 2 1 1 4 ... N N 実行制御部
- 2 1 1 5 ... N N 学習制御部
- 2 1 1 6 ... 縮約率補正部
- 2 1 2 0 ... 第 2 の結合情報群
- 2 2 0 0 ... 縮約後 N N
- 2 2 1 0 ... C P U
- 2 2 0 0 1 ... ネットワーク構造情報
- 3 0、3 1 ... データセット管理テーブル

請求の範囲

- [請求項1] 複数のニューロンと、前記ニューロン同士を紐付ける結合情報を複数含む結合情報群とを含むニューラルネットワークを格納する電子制御装置であって、
- 前記ニューラルネットワークは、
- 入力値が与えられたときに出力値を決定するための第1の前記結合情報群と、前記入力値として与えられた所定のデータに対する影響度が所定値を上回る複数の前記結合情報である第2の結合情報群とに基づき、間引き部が前記第1の結合情報群から少なくとも1つの結合情報を削除することで作成される第3の結合情報群を含むニューラルネットワークである電子制御装置。
- [請求項2] 請求項1に記載の電子制御装置において、
- 前記所定のデータは、車両走行に関する優先度が所定値を上回るデータである、電子制御装置。
- [請求項3] 請求項2に記載の電子制御装置において、
- 前記間引き部は、前記第2の結合情報群に含まれるそれぞれの前記結合情報に対応する操作係数の集合である操作係数群を決定し、それぞれの前記操作係数を前記第1の結合情報群に含まれる対応する前記結合情報に評価係数を算出して評価係数群とし、前記評価係数群のうち所定値を上回る前記評価係数について、対応する前記結合情報を前記第1の結合情報群の中から特定し、当該特定された結合情報以外の前記結合情報を前記第1の結合情報群から削除して、前記第3の結合情報群を作成する電子制御装置。
- [請求項4] 請求項3に記載の電子制御装置において、
- 係数特定部が、前記複数のニューロンのうち、前記所定のデータが与えられた場合の前記出力値が所定値を上回る場合に、当該出力値の決定に関与したニューロンを特定し、当該特定されたニューロンに紐付く結合情報を、前記第2の結合情報群を構成する前記結合情報とし

て特定する電子制御装置。

[請求項5]

請求項1に記載の電子制御装置において、

前記第3の結合情報群は、前記間引き部が前記第1の結合情報群から少なくとも1つの結合情報を削除し、ファインチューニングにより最適化された前記結合情報群である電子制御装置。

[請求項6]

サーバ、および電子制御装置を備え、複数のニューロンと、前記ニューロン同士を紐付ける結合情報を複数含む結合情報群とを含むニューラルネットワークを更新するニューラルネットワーク更新システムであって、

前記サーバは、記憶部と、演算部と、を備え、

前記記憶部は、入力値が与えられたときに出力値を決定するための第1の前記結合情報群を格納し、

前記演算部は、

所定のデータが前記入力値として与えられた場合に、前記所定のデータに対する影響度が所定値を上回る複数の前記結合情報である第2の結合情報群を特定する係数特定部と、

前記第1の結合情報群と前記第2の結合情報群とに基づき、前記第1の結合情報群から少なくとも1つの結合情報を削除して第3の結合情報群を作成する間引き部と、を備え、

前記電子制御装置は、前記第3の結合情報群を含むニューラルネットワークを格納するニューラルネットワーク更新システム。

[請求項7]

請求項6に記載のニューラルネットワーク更新システムにおいて、

前記所定のデータは、車両走行に関する優先度が所定値を上回るデータであるニューラルネットワーク更新システム。

[請求項8]

請求項7に記載のニューラルネットワーク更新システムにおいて、

前記間引き部は、前記第2の結合情報群に含まれるそれぞれの前記結合情報に対応する操作係数の集合である操作係数群を決定し、それぞれの前記操作係数を前記第1の結合情報群に含まれる対応する前記

結合情報に評価係数を算出して評価係数群とし、前記評価係数群のうち所定値を上回る前記評価係数について、対応する前記結合情報を前記第1の結合情報群の中から特定し、当該特定された結合情報以外の前記結合情報を前記第1の結合情報群から削除して、前記第3の結合情報群を作成するニューラルネットワーク更新システム。

[請求項9]

請求項8に記載のニューラルネットワーク更新システムにおいて、
前記係数特定部は、前記複数のニューロンのうち、前記所定のデータが与えられた場合の前記出力値が所定値を上回る場合に、当該出力値の決定に関与したニューロンを特定し、当該特定されたニューロンに紐付く結合情報を、前記第2の結合情報群を構成する前記結合情報として特定するニューラルネットワーク更新システム。

[請求項10]

請求項6に記載のニューラルネットワーク更新システムにおいて、
前記間引き部は、前記第1の結合情報群から少なくとも1つの結合情報を削除し、ファインチューニングにより前記結合情報群を最適化して前記第3の結合情報群を作成するニューラルネットワーク更新システム。

[図1]

図1

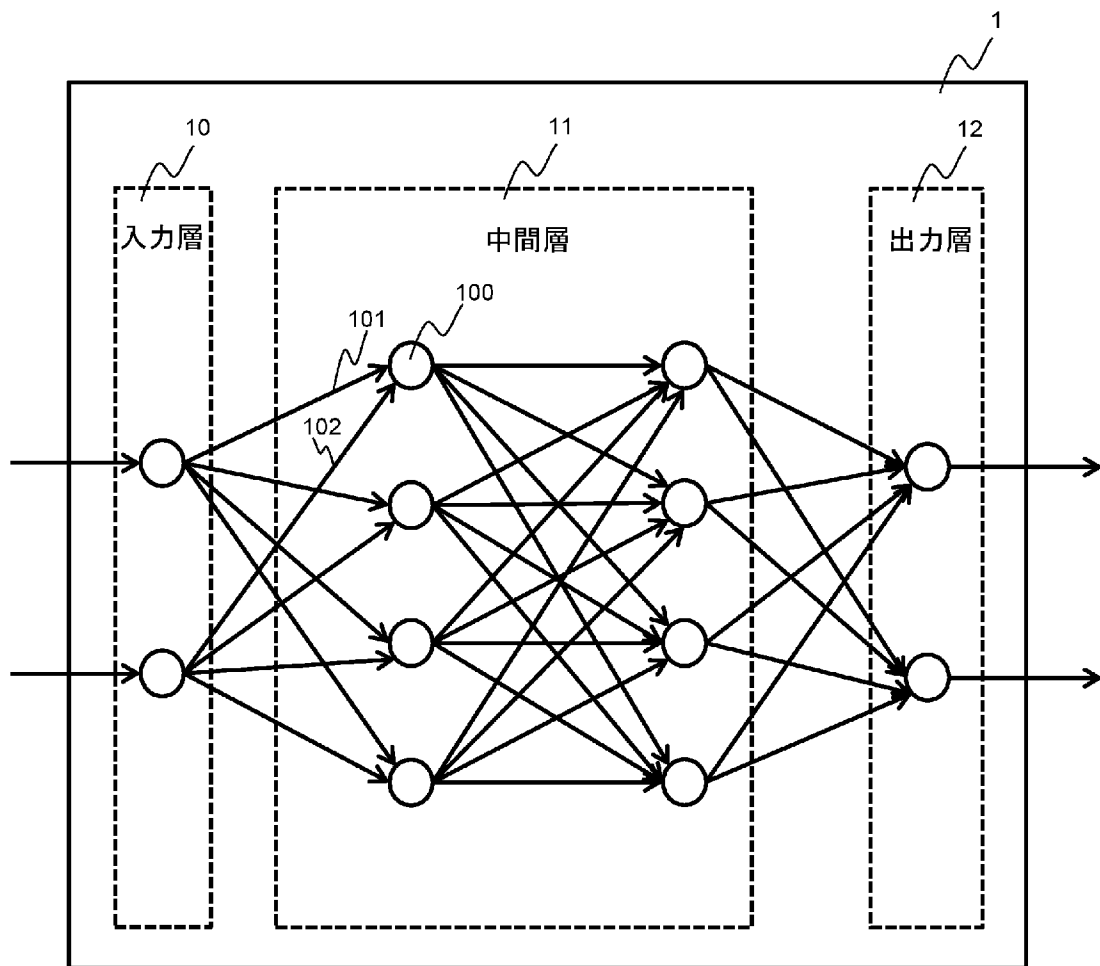
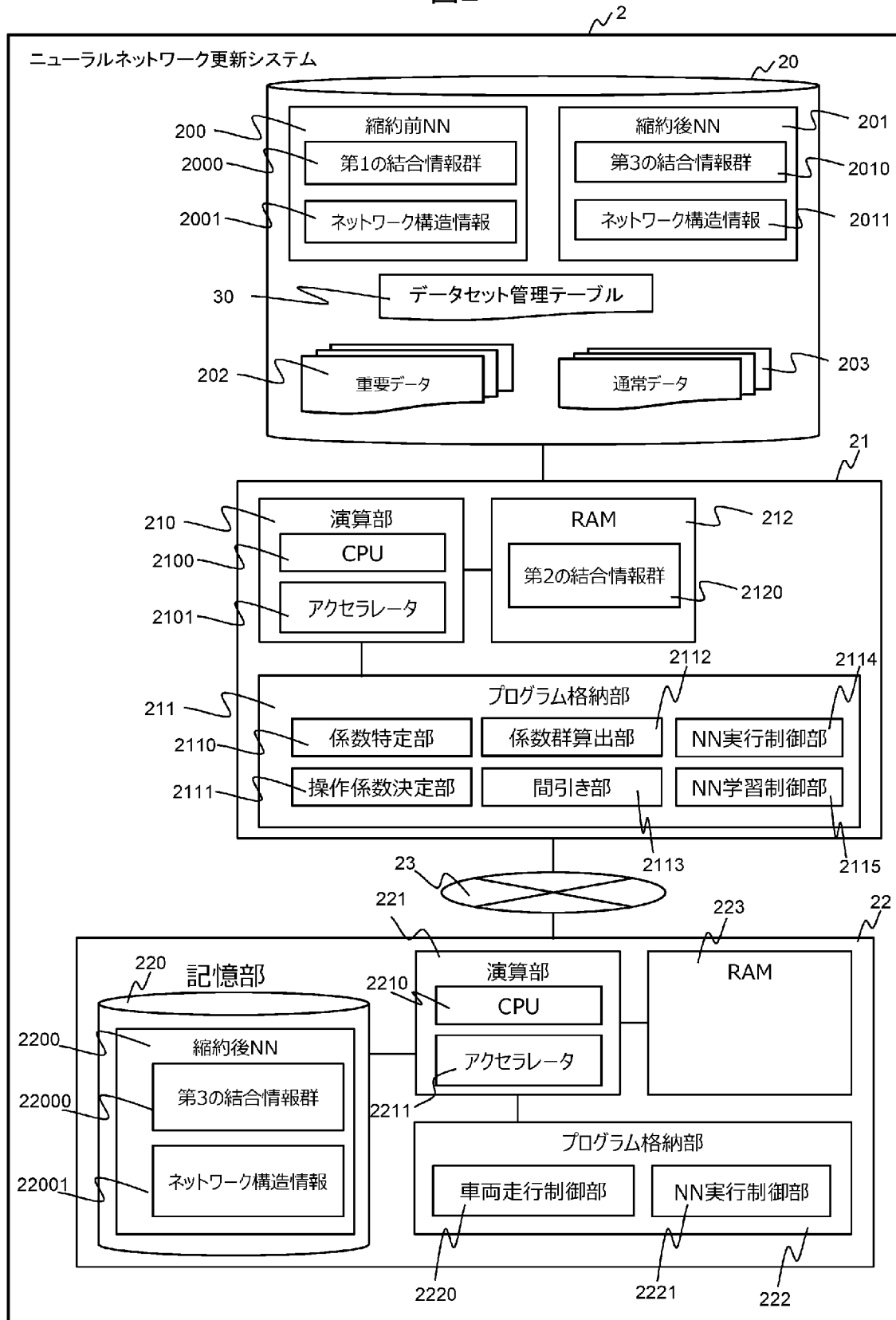


图2



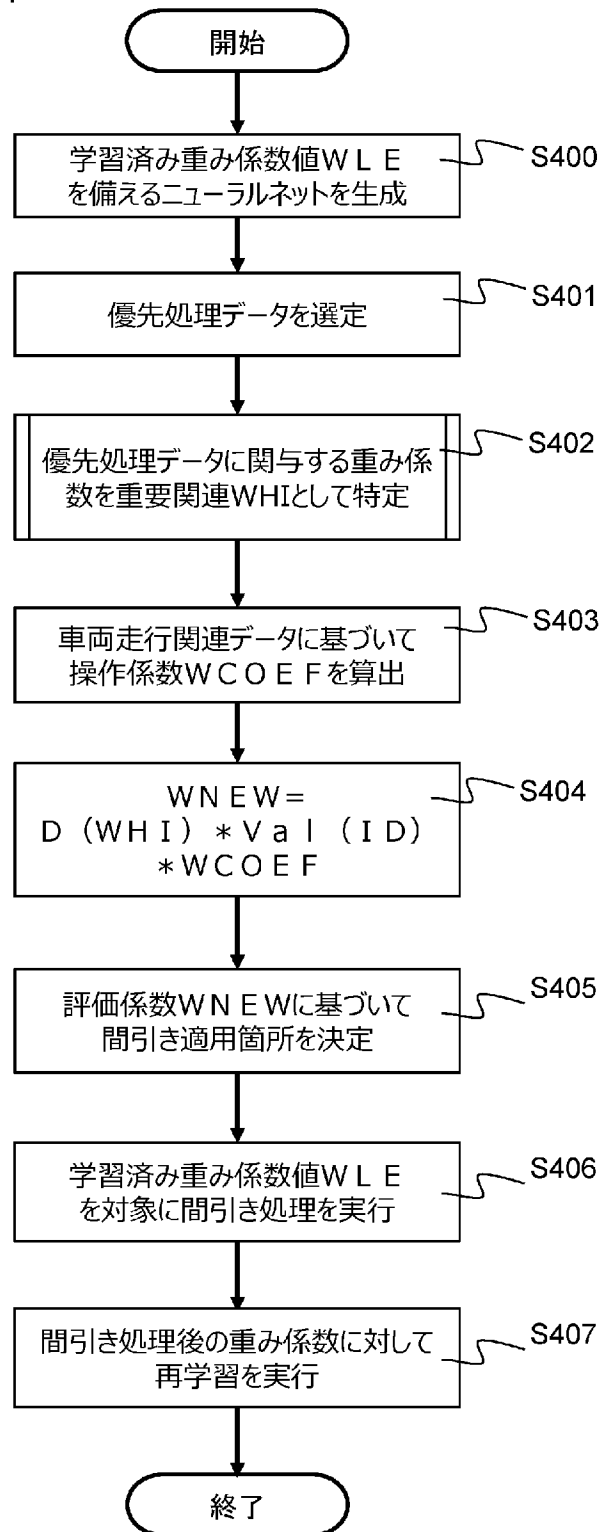
[図3]

図3

データ種別	優先度判定条件		優先度	操作係数
	相対距離 d	先行車フラグ		
オブジェクト検出データ1	$d < 30$	TRUE	3	2
オブジェクト検出データ2	$100 > d \geq 30$	TRUE	2	1.5
オブジェクト検出データ3	$d \geq 100$	TRUE	1	1
オブジェクト検出データ4	$d < 500$	FALSE	1	1

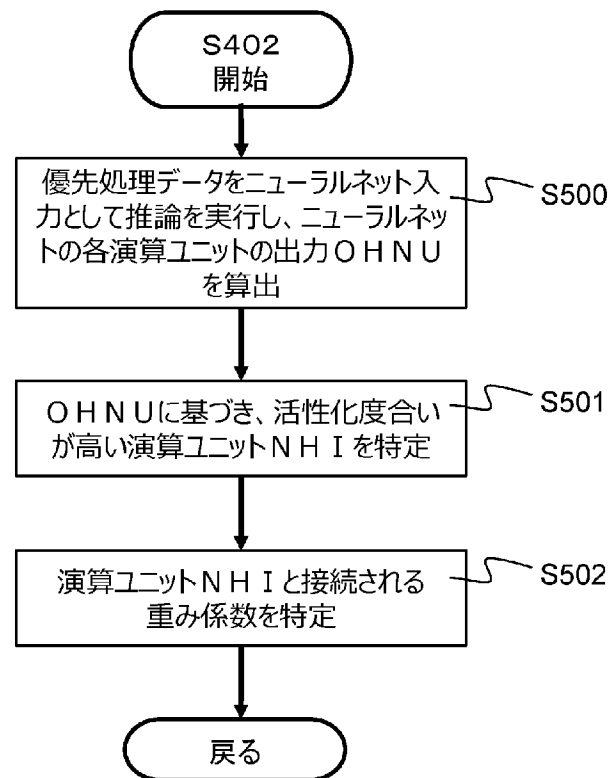
[図4]

図4



[図5]

図5



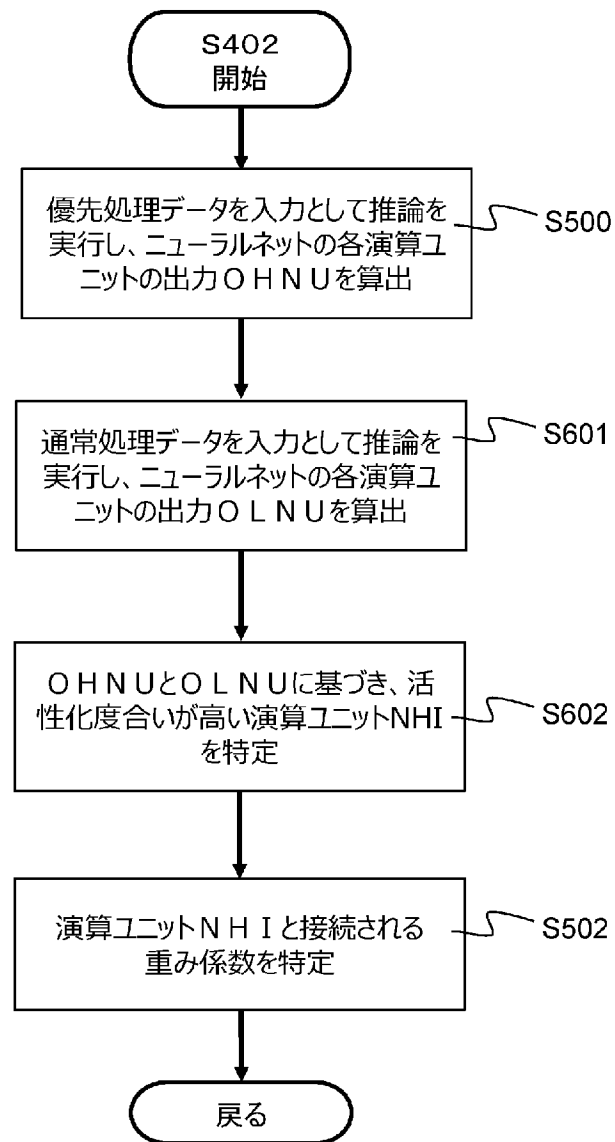
[図6]

図6

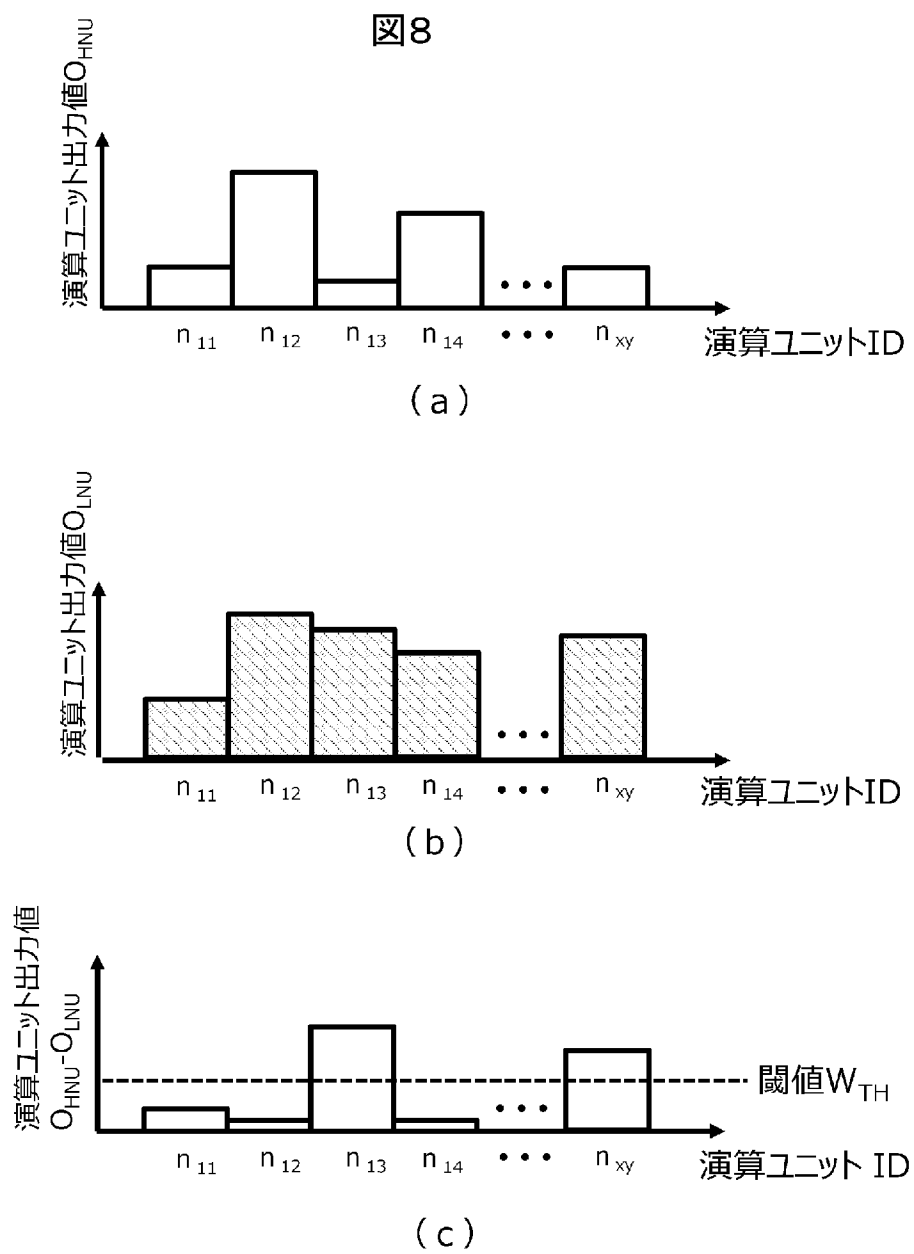
データ種別	優先度判定条件		優先度	重み係数
	識別クラス	オブジェクト高さ		
オブジェクト識別用データ1	車両、歩行者、自転車	—	3	2
オブジェクト識別用データ2	車両、歩行者、自転車以外の静止物体	20cm ≤ オブジェクト < 50cm	2	1.5
オブジェクト識別用データ3	車両、歩行者、自転車以外の静止物体	オブジェクト < 20cm	1	1

[図7]

図7

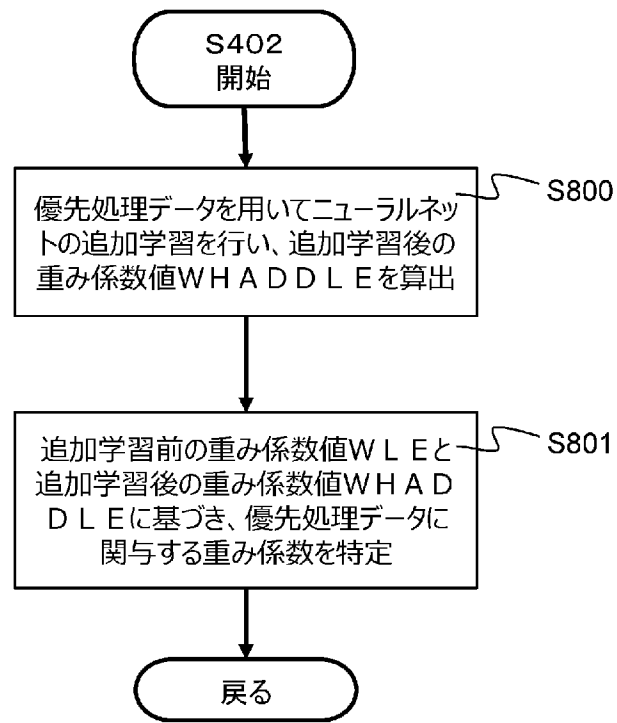


[図8]



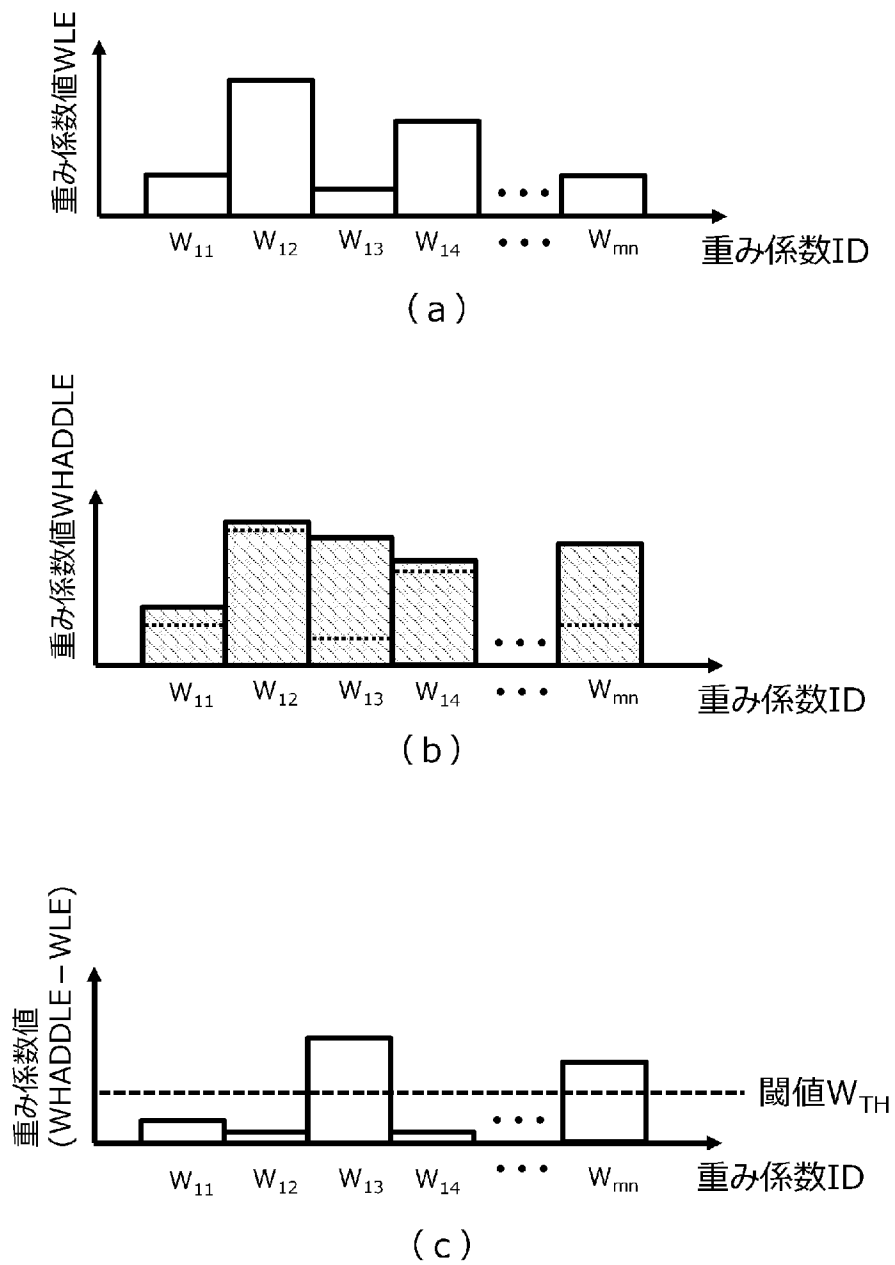
[図9]

図9



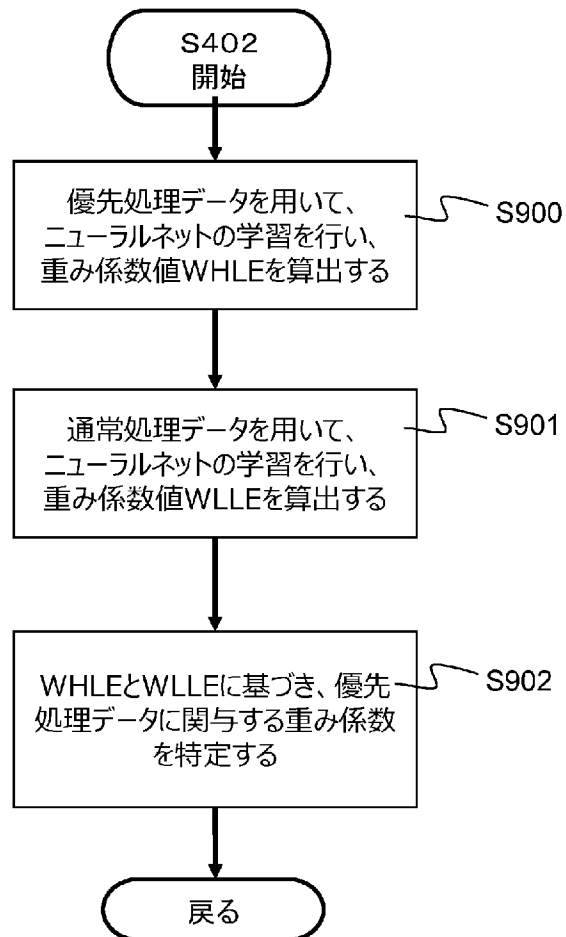
[図10]

図10



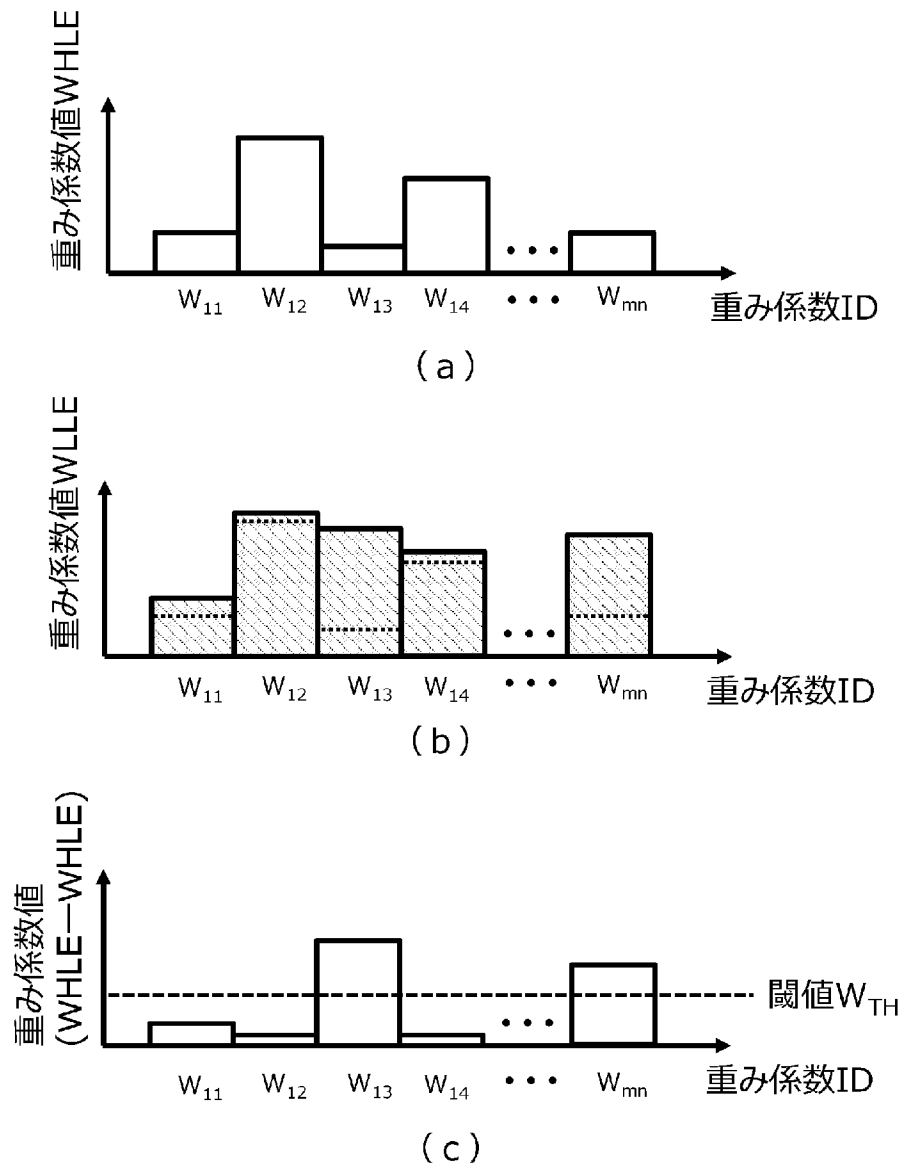
[図11]

図11



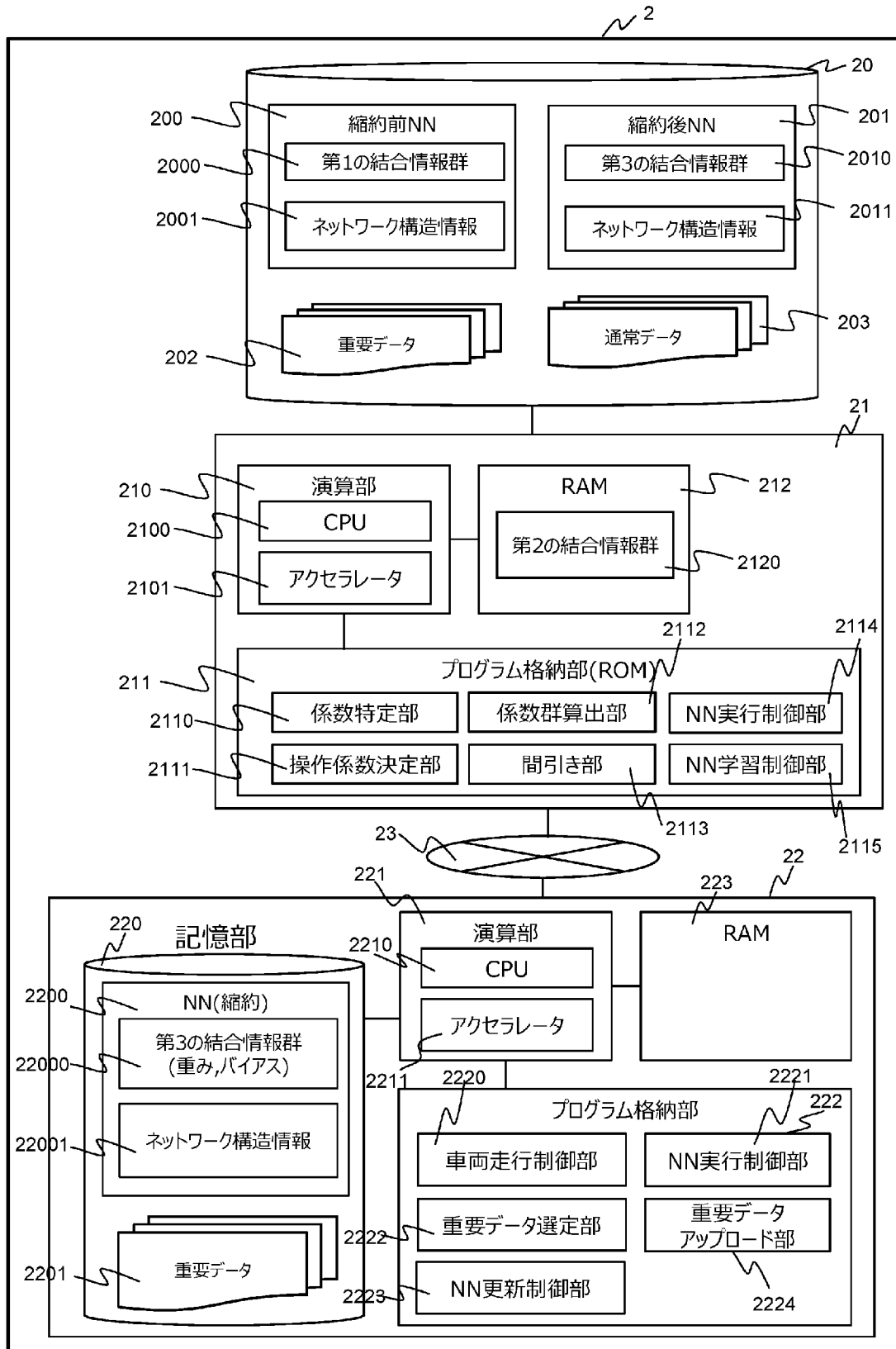
[図12]

図12



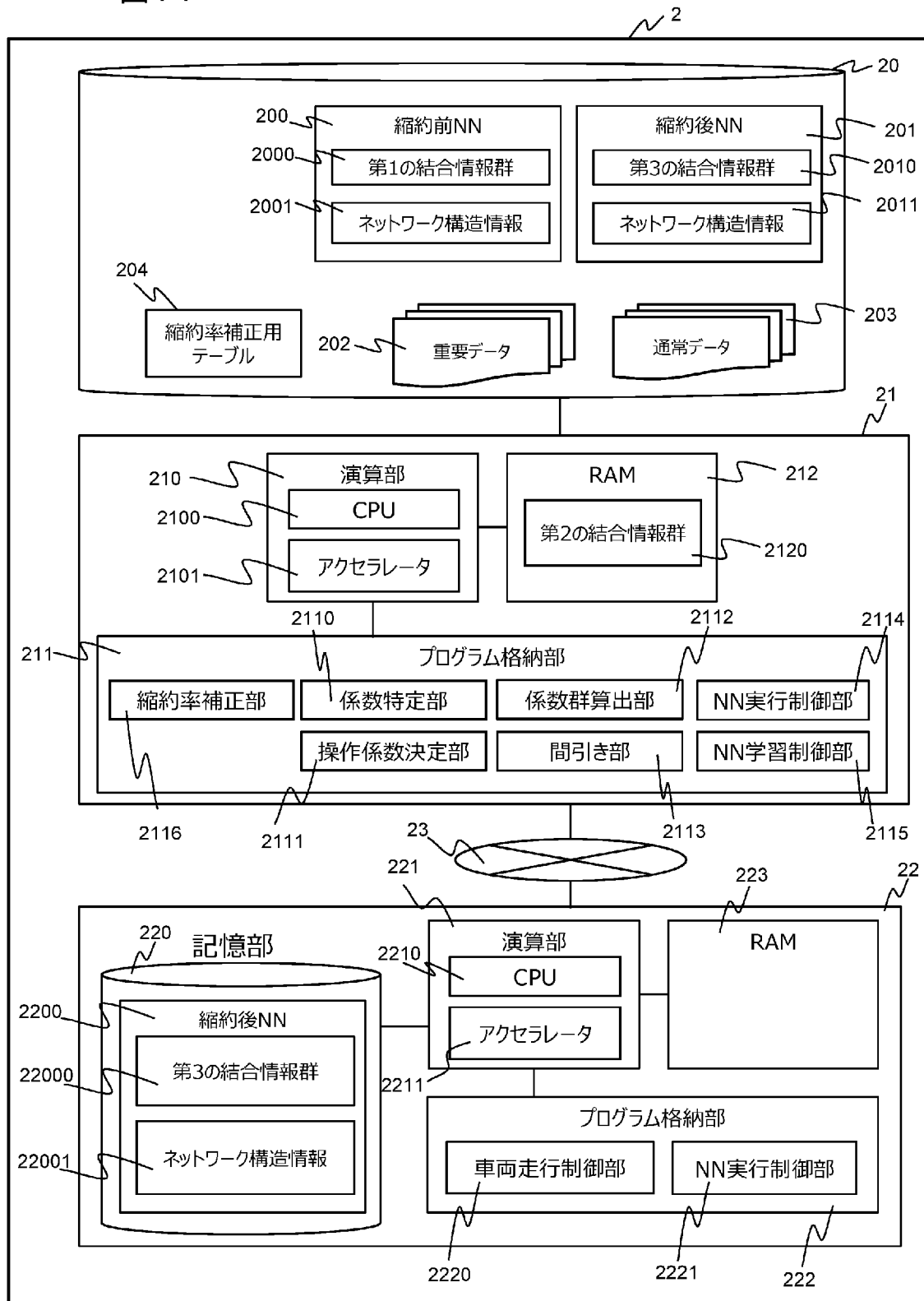
[図13]

図13



[図14]

図14



[図15]

図15

重要データに大きく関与する 演算ユニット数 u	縮約率
$0 \leq u < 20$	通常の縮約率
$20 \leq u < 40$	通常の縮約率/2
$20 \leq u < 40$	通常の縮約率/3
$40 \leq u$	通常の縮約率/4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/032837

A. CLASSIFICATION OF SUBJECT MATTER Int.Cl. G06N3/08 (2006.01) i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) Int.Cl. G06N3/02-3/10 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2019 Registered utility model specifications of Japan 1996-2019 Published registered utility model applications of Japan 1994-2019		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2018/0096249 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 05 April 2018, abstract, paragraphs [0005], [0029], [0041]-[0046], fig. 3-4 & KR 10-2018-0037558 A	1-10
A	JP 6-60051 A (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) 04 March 1994, abstract, claim 1 & US 5636327 A, abstract, claim 1 & EP 533193 A2	3, 8
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 18 October 2019 (18.10.2019)		Date of mailing of the international search report 05 November 2019 (05.11.2019)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan		Authorized officer Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06N3/08(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G06N3/02-3/10

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019年

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

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	US 2018/0096249 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2018.04.05, [要約], 段落[0005], [0029]及び [0041]-[0046]並びに[図3]-[図4] & KR 10-2018-0037558 A	1-10
A	JP 6-60051 A (松下電器産業株式会社) 1994.03.04, [要約]及び[請求項1] & US 5636327 A, [要約]及び[請求項1] & EP 533193 A2	3, 8

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

18.10.2019

国際調査報告の発送日

05.11.2019

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

多賀 実

5 B

1599

電話番号 03-3581-1101 内線 3545