

(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2015-95215

(P2015-95215A)

(43) 公開日 平成27年5月18日(2015.5.18)

| | | | | |
|--------------|-------------|------------------|-----------|-------------|
| (51) Int.Cl. | | F I | | テーマコード (参考) |
| GO6N | 3/04 | (2006.01) | GO6N 3/04 | F |
| GO6N | 3/08 | (2006.01) | GO6N 3/08 | Z |

審査請求 未請求 請求項の数 12 O L (全 29 頁)

| | | | |
|-----------|------------------------------|----------|--|
| (21) 出願番号 | 特願2013-235845 (P2013-235845) | (71) 出願人 | 502324066 株式会社デンソーアイティラボラトリ 東京都渋谷区渋谷2-15-1 渋谷クロ スタワー28F |
| (22) 出願日 | 平成25年11月14日(2013.11.14) | (71) 出願人 | 000004260 株式会社デンソー 愛知県刈谷市昭和町1丁目1番地 |
| | | (74) 代理人 | 230104019 弁護士 大野 聖二 |
| | | (74) 代理人 | 100106840 弁理士 森田 耕司 |
| | | (74) 代理人 | 100113549 弁理士 鈴木 守 |
| | | (74) 代理人 | 100131451 弁理士 津田 理 |

最終頁に続く

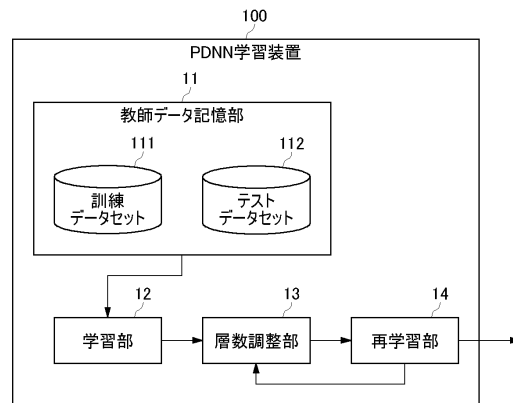
(54) 【発明の名称】 学習装置、学習プログラム、及び学習方法

(57) 【要約】

【課題】 多層ニューラルネットワークの層数を調整できる学習装置を提供する。

【解決手段】 PDNN学習装置100は、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、隠れ層に、出力層に順伝搬しないプローピングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部12と、プローピングニューロンのコストに基づいて多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とする層数調整部13と、層数調整部13にて上層を除去された多層ニューラルネットワークについて再学習を行う再学習部14とを備えている。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、

隠れ層に、出力層に順伝搬しないプロービングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、前記プロービングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプロービングニューロンを出力層とする層数調整部と、を備えたことを特徴とする学習装置。

【請求項 2】

前記層数調整部にて上層を除去された多層ニューラルネットワークについて再学習を行う再学習部をさらに備えたことを特徴とする請求項 1 に記載の学習装置。

【請求項 3】

前記層数調整部は、さらに、前記再学習部にて再学習された多層ニューラルネットワークについて、前記プロービングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプロービングニューロンを出力層とすることを特徴とする請求項 2 に記載の学習装置。

【請求項 4】

前記層数調整部は、コストが最小となる前記プロービングニューロンを前記出力層とすることを特徴とする請求項 1 ないし 3 のいずれか一項に記載の学習装置。

【請求項 5】

前記層数調整部は、前記多層ニューラルネットワークの出力層のコストが最小でない場合に、コストが最小である前記プロービングニューロンを有する層よりも上層の少なくとも 1 つの前記プロービングニューロンを残して前記多層ニューラルネットワークの上層を除去することを特徴とする請求項 3 に記載の学習装置。

【請求項 6】

前記プロービングニューロンに接続する部分が層構造を有することを特徴とする請求項 1 ないし 5 のいずれか一項に記載の学習装置。

【請求項 7】

教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、

請求項 1 ないし 6 のいずれか一項に記載の学習装置で学習された多層ニューラルネットワークから前記プロービングニューロンを除去した多層ニューラルネットワークを初期多層ニューラルネットワークとして使用した学習を行うことを特徴とする学習装置。

【請求項 8】

請求項 1 ないし 6 のいずれか一項の学習装置によって得られた多層ニューラルネットワークを用いて識別を行う識別器であって、

前記プロービングニューロンによって識別を行うことで早期棄却を行うことを特徴とする識別器。

【請求項 9】

前記プロービングニューロンによって出力層のいずれかのユニットに該当するか否かを判断することで前記早期棄却を行うことを特徴とする請求項 8 に記載の識別器。

【請求項 10】

教師あり学習によって順伝搬型の畳み込みニューラルネットワークの学習を行う学習装置であって、

隠れ層に、マップに連結するプロービングニューロンを有する畳み込みニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、

前記プロービングニューロンのコストに基づいて前記畳み込みニューラルネットワークの上層を除去して、残った最上層のプロービングニューロンを畳み込みの出力層とする層数調整部と、

10

20

30

40

50

を備えたことを特徴とする学習装置。

【請求項 1 1】

コンピュータを、

教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、

隠れ層に、出力層に順伝搬しないプローピングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、

前記プローピングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とする層数調整部と、
を備えた学習装置として機能させることを特徴とする学習プログラム。

10

【請求項 1 2】

教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習方法であって、

隠れ層に、出力層に順伝搬しないプローピングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習ステップと、

前記プローピングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とする層数最適化ステップと、

を含むことを特徴とする学習方法。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置、学習プログラム、及び学習方法に関するものである。

【背景技術】

【0002】

教師あり学習によって得られる多層ニューラルネットワークは、識別能力が高く優れた計算モデルである。教師あり学習によって多層ニューラルネットワークの構造を決定するに際して、ある与えられた訓練データ（教師データともいう。）に対してどのような構造の多層ニューラルネットワークが高い汎化能力（学習に使用していないデータに対する識別性能）を与えるかを説明する理論はいまだ確立されておらず、従来より、汎化能力の高い多層ニューラルネットワークの構造を得るための手法として、発見的手法が採用されてきた。

30

【0003】

例えば、X. Liang, "Removal of Hidden Neurons by Crosswise Propagation", *Neural Information Processing-Letters and Reviews*, Vol. 6, No 3, 2005.（非特許文献1）では、多層ニューラルネットワークの各隠れ層のユニットを1つずつ除去することで最適なネットワーク構造を構築する方法が提案されている。この方法は、初期設定された多層ニューラルネットワーク（以下単に「初期多層ニューラルネットワーク」又は「初期ネットワーク」という。）を十分に訓練した状態で、以下の要領に従ってユニットを削減する：訓練データに対し、同じ層の異なるユニットの出力同士の相関を計算する；相関が最も高いものを1つ除去する；ユニットの除去の後、それ以外の重みの学習を再開する；再学習とユニットの除去をコスト関数が上昇に転じるまで繰り返す。なお、最初のネットワーク構造は手動で与えられる。

40

【0004】

また、特許第3757722号公報（特許文献1）には、教師あり学習における多層ニューラルネットワークの中間層（隠れ層）のユニット数を最適化する方法が記載されている。

【先行技術文献】

50

【特許文献】

【0005】

【特許文献1】特許第3757722号公報

【非特許文献】

【0006】

【非特許文献1】X. Liang, "Removal of Hidden Neurons by Crosswise Propagation", Neural Information Processing-Letters and Reviews, Vol. 6, No 3, 2005.

【非特許文献2】Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network", Advances in Neural Information Processing Systems (NIPS), pp. 396-404, 1990. 10

【発明の概要】

【発明が解決しようとする課題】

【0007】

多層ニューラルネットワークの構造を決定するための上記の従来手法に共通することは、初期ネットワークを最初に訓練し、汎化能力が良くなると期待できる指標に従ってユニットを増減することで隠れ層のユニット数を自動決定する点である。即ち、従来手法は、層数を固定してユニット数を最適化するものであって、層数自体を最適化するものではない。

【0008】

多層ニューラルネットワークは一般に識別性能が良いとされるが、多層になればなるほど識別に時間がかかり、従って層数は識別性能と計算量を大きく左右するパラメタである。しかしながら、上述のように、従来、層数の最適な決定方法は提案されていない。また、畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)においても、フィルタ層の層数、またフィルタ層に続く全結合層の層数は、設計者によって手動で決められるのが現状であり、フィルタ層と全結合層の最適な層数の決定方法は提案されていない。 20

【0009】

本発明は、上記の問題に鑑みてなされたものであり、多層ニューラルネットワークの層数を自動で調整できる学習装置、学習プログラム、及び学習方法を提供することを目的とする。 30

【課題を解決するための手段】

【0010】

本発明の第1の態様の学習装置は、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、隠れ層に、出力層に順伝搬しないプロベイングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、前記プロベイングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプロベイングニューロンを出力層とする層数調整部とを備えたことを特徴とする。

【0011】

この構成によれば、層数調整部が学習部における学習に基づいて、多層ニューラルネットワークの不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できる。 40

【0012】

本発明の第2の態様の学習装置は、第1の態様の学習装置において、前記層数調整部にて上層を除去された多層ニューラルネットワークについて再学習を行う再学習部をさらに備えたことを特徴とする。

【0013】

この構成によれば、不要な上層を除去した後に再学習を行うので、コストをさらに低減 50

できる。

【0014】

本発明の第3の態様の学習装置は、第2の態様の学習装置において、前記層数調整部は、さらに、前記再学習部にて再学習された多層ニューラルネットワークについて、前記プローピングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とすることを特徴とする。

【0015】

この構成によれば、再学習部による再学習の後に再度上層の除去を試み、必要に応じて上層を除去するので、さらに層数を最適化できる。

【0016】

本発明の第4の態様の学習装置は、第1ないし第3のいずれかの態様の学習装置において、前記層数調整部は、コストが最小となる前記プローピングニューロンを前記出力層とすることを特徴とする。

【0017】

この構成によれば、出力層のコストを最小化して、層数を最適化できる。

【0018】

本発明の第5の態様の学習装置は、第3の態様の学習装置において、前記層数調整部は、前記多層ニューラルネットワークの隠れ層の最上層の前記プローピングニューロンのコストが最小でない場合に、コストが最小である前記プローピングニューロンを有する層よりも上層の少なくとも1つの前記プローピングニューロンを残して前記多層ニューラルネットワークの上層を除去することを特徴とする。

【0019】

この構成によれば、層を上層側から徐々に除去していくので、複数回の再学習によって、最適な層数を決定することができる。

【0020】

本発明の第6の態様の学習装置は、第1ないし第5のいずれかの態様の学習装置において、前記プローピングニューロンに接続する部分が層構造を有することを特徴とする。

【0021】

この構成によれば、プローピングニューロンに接続する部分が層構造をなすことで、プローピングニューロンの識別性能を高めることができ、さらに、隠れ層の線形分離性能の最適化の制約がないので、出力層における識別性能の向上も期待できる。

【0022】

本発明の第7の態様の学習装置は、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、第1ないし第6のいずれかの態様の学習装置で学習された多層ニューラルネットワークから前記プローピングニューロンを除去した多層ニューラルネットワークを初期多層ニューラルネットワークとして使用した学習を行うことを特徴とする。

【0023】

この構成によれば、識別性能の高い初期多層ニューラルネットワークを提供できる。

【0024】

本発明の第8の態様の識別器は、第1ないし第6のいずれかの学習装置によって得られた多層ニューラルネットワークを用いて識別を行う識別器であって、前記プローピングニューロンによって識別を行うことで早期棄却を行うことを特徴とする。

【0025】

この構成によれば、早期棄却を採用した高速な識別が可能となる。

【0026】

本発明の第9の態様の識別器は、第8の態様の識別器において、前記プローピングニューロンによって出力層のいずれかのユニットに該当するか否かを判断することで前記早期棄却を行うことを特徴とする。

【0027】

10

20

30

40

50

この構成によれば、出力層のいずれかのユニットに該当するか否かを早期に判断して、計算量を削減できる。

【0028】

本発明の第10の態様の学習装置は、教師あり学習によって順伝搬型の畳み込みニューラルネットワークの学習を行う学習装置であって、隠れ層に、マップに連結するプローピングニューロンを有する畳み込みニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、前記プローピングニューロンのコストに基づいて前記畳み込みニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを畳み込みの出力層とする層数調整部とを備えたことを特徴とする。

【0029】

この構成によれば、層数調整部が学習部における学習に基づいて、畳み込みニューラルネットワークの不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する畳み込みニューラルネットワークを取得できる。

【0030】

本発明の第11の態様の学習プログラムは、コンピュータを、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置であって、隠れ層に、出力層に順伝搬しないプローピングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習部と、前記プローピングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とする層数調整部とを備えた学習装置として機能させることを特徴とする。

【0031】

この構成によっても、層数調整部が学習部における学習に基づいて、不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できる。

【0032】

本発明の第12の態様の学習方法は、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習方法であって、隠れ層に、出力層に順伝搬しないプローピングニューロンを有する多層ニューラルネットワークについて、訓練データセットを用いたコスト計算による学習を行う学習ステップと、前記プローピングニューロンのコストに基づいて前記多層ニューラルネットワークの上層を除去して、残った最上層のプローピングニューロンを出力層とする層数最適化ステップとを含むことを特徴とする。

【0033】

この構成によっても、学習ステップにおける学習に基づいて、層数最適化ステップにて不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できる。

【発明の効果】

【0034】

本発明によれば、層数調整部が学習部における学習に基づいて、不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できる。

【図面の簡単な説明】

【0035】

【図1】本発明の実施の形態における多層ニューラルネットワーク学習装置の構成を示すブロック図

【図2】本発明の実施の形態における初期多層ニューラルネットワークの構成を示す図

【図3】従来の多層ニューラルネットワークの構造を示す図

【図4】本発明の実施の形態における初期多層ニューラルネットワークの構造を示す図

【図5】本発明の実施の形態における層数と学習終了時のテストコストとの関係を示す図

10

20

30

40

50

【図 6】本発明の実施の形態における層を除去した後の多層ニューラルネットワークを示す図

【図 7】本発明の実施の形態における再学習部による処理を説明する図

【図 8】本発明の実施の形態における多層ニューラルネットワーク学習装置による学習方法のフロー図

【図 9】本発明の実施の形態の変形例 1 における多層ニューラルネットワーク学習装置における P D N N 学習方法のフロー図

【図 10】本発明の実施の形態の変形例 2 における多層ニューラルネットワークの構造を示す図

【図 11】本発明の実施の形態の応用例 1 における多層ニューラルネットワーク学習装置の構成を示すブロック図

【図 12】本発明の実施の形態におけるファインチューニングされた多層ニューラルネットワークからプローピングニューロンを除去した多層ニューラルネットワークのコストを示す図

【図 13】本発明の実施の形態の応用例 2 における車載カメラを使って歩行者を検知する例を示す図

【図 14】本発明の実施の形態の応用例 2 における早期棄却の処理を説明する図

【図 15】本発明の実施の形態に 応用例 3 における早期棄却の処理を説明する図

【図 16】本発明の実施の形態の応用例 4 で用いる手書き数字のデータセットを示す図

【図 17】本発明の実施の形態の応用例 4 における多層ニューラルネットワークの構造を示す図

【図 18】従来の学習によって得られた多層ニューラルネットワークの構造を示す図

【図 19】図 16 のデータセットを用いてテストをした結果を示すグラフ

【図 20】従来の畳み込みニューラルネットワークの構造を示す図

【図 21】本発明の実施の形態の応用例 6 における畳み込みニューラルネットワークの構造を示す図

【発明を実施するための形態】

【0036】

以下、本発明の実施の形態の多層ニューラルネットワーク学習装置について、図面を参照しながら説明する。なお、以下に説明する実施の形態は、本発明を実施する場合の一例を示すものであって、本発明を以下に説明する具体的構成に限定するものではない。本発明の実施にあたっては、実施の形態に応じた具体的構成が適宜採用されてよい。

【0037】

図 1 は、本実施の形態の多層ニューラルネットワーク学習装置の構成を示すブロック図である。本実施の形態の多層ニューラルネットワーク学習装置は、後述するようにプローピングニューロンが追加された多層ニューラルネットワークの学習装置であり、以下では、プローピングニューロンが追加された多層ニューラルネットワークをプローブ深層ニューラルネットワーク（P D N N : Probed Deep Neural Network）と呼び、その学習装置を P D N N 学習装置と呼ぶこととする。また、以下では、多層ニューラルネットワークを単に「ネットワーク」ともいう。

【0038】

P D N N 学習装置 100 は、教師データ記憶部 11 と、学習部 12 と、層数調整部 13 と、再学習部 14 とを備えている。P D N N 学習装置 100 は、補助記憶装置（H D D、S S D 等）、主記憶装置（R A M 等）、処理装置（C P U 等）、入出力装置（各種入出力ポート及び通信モジュール）等を備えたコンピュータによって実現される。特に、教師データ記憶部 11 は、補助記憶装置によって実現され、学習部 12、層数調整部 13、及び再学習部 14 は、コンピュータが本実施の形態の学習プログラムを実行することによって実現される。

【0039】

学習部 12 は、与えられた初期多層ニューラルネットワークについてコストを計算する

10

20

30

40

50

ことで学習を行う。層数調整部 13 は、学習部 12 にて計算された初期多層ニューラルネットワークのコストに基づいて、初期多層ニューラルネットワークの上層を除去することで、多層ニューラルネットワークの層数を調整して最適化する。再学習部 14 は、層数調整部 13 にて層数が最適化された多層ニューラルネットワークについて再学習を行って、多層ニューラルネットワークをチューニングする。層数調整部 13 は、さらに、再学習部 14 で調整された多層ニューラルネットワークについても、必要に応じて上層の除去を行うことで、再度層数を調整して最適化する。

【0040】

教師データ記憶部 11 には、教師データとして、ネットワークの入力となる多次元ベクトルと、それに対応した、ネットワークの出力となるターゲット値（多次元ベクトルまたはスカラー）の組が多数記憶されている。教師データは、ネットワークを学習するために使用される訓練データセット 111 と、学習されたネットワークを評価するために使用されるテストデータセット 112 に分割されて教師データ記憶部 11 に記憶されている。

10

【0041】

PDNN 学習装置 100 の各部における処理の説明の前に、まず、本発明の実施の形態において対象とする多層ニューラルネットワークの種類について説明する。本実施の形態の多層ニューラルネットワークは、層を跨いだユニット間の結合はないものとする。また、ある層の全ユニットと、その次の層の全ユニットとは、全てが結合していてもよいし、一部が結合していなくてもよい。対象とするネットワークは、入力層から出力層まで順番に信号が伝搬する順伝搬（フィードフォワード）型とする。

20

【0042】

図 2 は、従来の多層ニューラルネットワークとの比較において、本実施の形態の初期多層ニューラルネットワークの構成を示す図である。図 2 の上段が本実施の形態の初期多層ニューラルネットワークの一例であり、下段が従来の多層ニューラルネットワークである。図 2 に示すように、本実施の形態の初期多層ニューラルネットワークは、従来の多層ニューラルネットワークに、プロービングニューロン（PN: Probing Neuron）というユニットを追加したネットワーク構造を有する。PN は、ターゲット値に近づくように学習された、隠れ層に属するユニットである。

【0043】

以下では、まず、従来の多層ニューラルネットワークの学習方法を説明し、次いで本実施の形態の PDNN 学習装置 100 による PDNN の学習方法を説明する。図 3 は、従来の多層ニューラルネットワークの構造を示す図である。図 3 において、層数 L と各層のユニット数 D_l ($l = 1, 2, \dots, L$) はあらかじめ与えられているものとする。第 n ($n = 1, 2, \dots, N$) 番目の入力ベクトルを下式 (1)、式 (2) のように定義する。

30

【数 1】

$$x_0^n \in \mathbb{R}^{D_0} \quad (1)$$

【数 2】

$$\bar{x}_0^n = [1|x_0^{nT}]^T \quad (2)$$

40

【0044】

すべての $l = 1, 2, \dots, L$ に対して、下式 (3) ~ (7) の変数を定める。

【数 3】

$$W_l \in \mathbb{R}^{D_l \times D_{l-1}} \quad (3)$$

【数 4】

$$\bar{W}_l = [b_l | W_l], b_l \in \mathbb{R}^{D_l} \quad (4)$$

【数 5】

$$a_l^n = \bar{W}_l \bar{x}_{l-1}^n \quad (5)$$

10

【数 6】

$$\begin{aligned} x_l^n &= h(a_l^n) \quad \forall l = 1, 2, \dots, L-1 \\ x_L^n &= a_L^n \quad (6) \end{aligned}$$

【数 7】

$$\bar{x}_l^n = [1 | x_l^{nT}]^T \quad (7)$$

20

なお、 h は単調増加かつ微分可能な非線形活性化関数である。

【0045】

最適化問題として、下式(8)を最小化する

【数 8】

$$\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L$$

を求める。

【数 9】

30

$$E(\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L) = \sum_{n=1}^N E^n(\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L) \quad (8)$$

ここで、

【数 10】

$$E^n = \frac{1}{2} \|x_L^n - t^n\|_2^2$$

40

であり、 t^n は、次元 D_L の既知の正解出力ベクトル(スカラー)である。なお、 E^n は、真値との差の二乗の総和の他、クロスエントロピーを採用することもできる。

【0046】

学習は、出力層に近い重みから順に勾配法による更新を繰り返す誤差逆伝搬法を用いる。なお、この時の初期値は乱数によって与えられる。勾配法は下式(9)で表現される。

【数 1 1】

$$\bar{W}_l \rightarrow \bar{W}_l - \epsilon \frac{\partial E}{\partial \bar{W}_l}, \quad \epsilon \ll 1 \quad (9)$$

勾配法で使用する微分の計算方法は以下のとおりである。なお、以下の式において下付きの i 、 j 、 k は、ベクトルや行列のインデクスである。

【0047】

まず、最終の重み層の微分は、下式(10)～(12)の計算で求められる。

【数 1 2】

10

$$\frac{\partial E_L^n}{\partial \bar{W}_{Lij}} = \sum_{k=1}^{D_L} \Delta_L^n{}_k \frac{\partial a_{Lk}^n}{\partial \bar{W}_{Lij}} \quad (10)$$

【数 1 3】

$$\left\{ \begin{array}{l} \frac{\partial a_{Lk}^n}{\partial \bar{W}_{Lij}} = \delta_{ki} \bar{x}_{L-1}^n{}_j \quad (11) \\ \Delta_L^n{}_k \equiv \frac{\partial E^n}{\partial a_{Lk}^n} = (a_L^n - t^n)_k \quad (12) \end{array} \right.$$

20

【0048】

$l = D - 1, D - 2, \dots, 1$ の重み層の微分は、下式(13)～(16)の計算で求められる。

【数 1 4】

30

$$\frac{\partial E^n}{\partial \bar{W}_{lij}} = \sum_{k=1}^{D_l} \Delta_l^n{}_k \frac{\partial a_{lk}^n}{\partial \bar{W}_{lij}} \quad (13)$$

【数 1 5】

$$\left\{ \begin{array}{l} \frac{\partial a_{lk}^n}{\partial \bar{W}_{lij}} = \delta_{ki} \bar{x}_{l-1}^n{}_j \quad (14) \\ \Delta_l^n{}_k \equiv \frac{\partial E^n}{\partial a_{lk}^n} = \sum_{j=1}^{D_{l+1}} \frac{\partial E^n}{\partial a_{l+1}^n{}_j} \frac{\partial a_{l+1}^n{}_j}{\partial a_{l+1}^n{}_k} = \sum_{j=1}^{D_{l+1}} \Delta_{l+1}^n{}_j \frac{\partial a_{l+1}^n{}_j}{\partial a_{l+1}^n{}_k} \quad (15) \\ \frac{\partial a_{l+1}^n{}_j}{\partial a_{l+1}^n{}_k} = W_{l+1,jk} h'(a_{l+1}^n{}_k) \quad (16) \end{array} \right.$$

40

【0049】

50

次に、本実施の形態のPDNNの学習方法を説明する。PDNNの学習方法では、まず、十分な層数を持ったPDNNが初期ネットワークとして設定される。そして、PDNN学習装置100は、以下の手順によって最適な層数を決定する。

【0050】

図4は、PDNNの初期ネットワークの構造を示す図である。層数Lと各層のユニット数 $D_l = (l = 1, 2, \dots, L)$ はあらかじめ与えられるものとする。また、各隠れ層に置かれるPNの個数は出力ユニット数と同数である。訓練データセット111の第n ($n = 1, 2, \dots, N$)番目の入力ベクトルを下式(17)、(18)のように定義する。

【数16】

$$x_0^n \in \mathbb{R}^{D_0} \quad (17)$$

10

【数17】

$$\bar{x}_0^n = [1|x_0^{nT}]^T \quad (18)$$

【0051】

$l = 1$ に対して、下式(19)~(23)の変数を定める。

【数18】

$$W_1 \in \mathbb{R}^{D_1 \times D_0} \quad (19)$$

20

【数19】

$$\bar{W}_1 = [b_1|W_1], b_1 \in \mathbb{R}^{D_1} \quad (20)$$

【数20】

$$a_1^n = \bar{W}_1 \bar{x}_0^n \quad (21)$$

30

【数21】

$$x_1^n = h(a_1^n) \quad (22)$$

【数22】

$$\bar{x}_1^n = [1|x_1^{nT}]^T \quad (23)$$

【0052】

また、 $l = 2, 3, \dots, L - 1$ に対して、下式(24)~(28)の変数を定める。

【数23】

$$W_l \in \mathbb{R}^{(D_l + D_L) \times D_{l-1}} \quad (24)$$

【数24】

$$\bar{W}_l = [b_l|W_l], b_l \in \mathbb{R}^{D_l + D_L} \quad (25)$$

40

【数 2 5】

$$a_l^n_i = \sum_{j=1}^{D_{l-1}} \bar{W}_{lij} \bar{x}_{l-1}^n_j \quad (26)$$

【数 2 6】

$$x_l^n = h(a_l^n) \quad (27)$$

10

【数 2 7】

$$\bar{x}_l^n = [1|x_l^{nT}]^T \quad (28)$$

【0 0 5 3】

最終層 ($l = L$) の変数は次のように定める。

【数 2 8】

$$W_L \in \mathbb{R}^{D_L \times D_{L-1}} \quad (29)$$

20

【数 2 9】

$$\bar{W}_L = [b_L|W_L], \quad b_L \in \mathbb{R}^{D_L} \quad (30)$$

【数 3 0】

$$a_L^n_i = \sum_{j=1}^{D_{L-1}} \bar{W}_{Lij} \bar{x}_{L-1}^n_j \quad (31)$$

30

【数 3 1】

$$x_L^n = h(a_L^n) \quad (32)$$

【0 0 5 4】

学習部 1 2 は、最適化問題として、下式 (3 3) を最小化する

【数 3 2】

$$\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L$$

40

を求める。

【数 3 3】

$$\begin{aligned}
E(\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L) &= \sum_{n=1}^N E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[1,D_2]}, \dots, \bar{W}_L \right) \\
&+ E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[1,D_2]}, \dots, \bar{W}_{L-1[D_{L-1}+1, D_{L-1}+D_L]} \right) \\
&+ E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[1,D_2]}, \dots, \bar{W}_{L-2[D_{L-2}+1, D_{L-2}+D_L]} \right) \\
&\dots + E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[D_2+1, D_2+D_L]} \right) \quad (33)
\end{aligned}$$

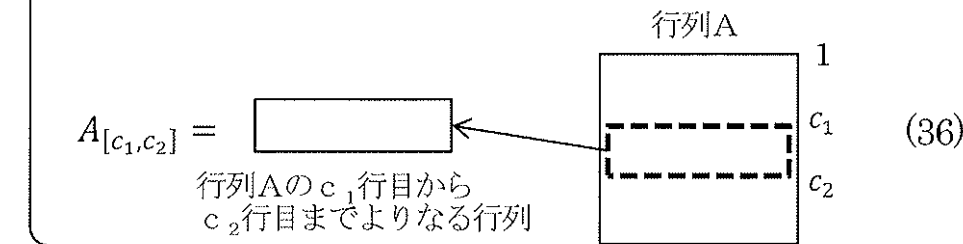
10

【数 3 4】

$$E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[1,D_2]}, \dots, \bar{W}_L \right) = \frac{1}{2} \sum_{i=1}^{D_L} (x_L^n{}_i - t^n{}_i)^2 \quad (34)$$

20

$$\begin{aligned}
E^n \left(\bar{W}_{1[1,D_1]}, \bar{W}_{2[1,D_2]}, \dots, \bar{W}_{l[D_l+1, D_l+D_L]} \right) &= \frac{1}{2} \sum_{i=1}^{D_L} (x_l^n{}_{D_l+i} - t^n{}_i)^2 \quad (35) \\
\forall l &= 1, 2, \dots, L-1
\end{aligned}$$



30

ここで、式(33)の右辺の第1項は出力層(第L層)のコストであり、右辺第2項は第L-1層のPNコストであり、右辺第3項は第L-2層のPNコストであり、右辺の最終項は第2層のPNコストである。

【0055】

なお、隠れ層においては、PNと真値との差の二乗の総和をPNコストとして採用してよく、出力層においては通常のニューラルネットワークと同じであり、二乗和の代わりにクロスエントロピーをコストとして採用してもよい。

40

【0056】

以下では、下式(37)、(38)のように標記を簡略化する。

【数 3 5】

$$E_l^n = E^n \left(\bar{W}_1, \bar{W}_2, \dots, \bar{W}_{l[D_l+1, D_l+D_L]} \right), \quad l = 2, 3, \dots, L-1 \quad (37)$$

【数 3 6】

$$E_L^n = E^n(\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L) \quad (38)$$

【0057】

学習は、通常のニューラルネットワークと同様に、勾配法による重みの更新則を適用して行う。まず、最終 L の重み層の微分は、下式 (39) ~ (41) の計算で求められる。

【数 3 7】

$$\frac{\partial E_L^n}{\partial \bar{W}_{Lij}} = \sum_{k=1}^{D_L} \Delta_L^n{}_k \frac{\partial a_L^n{}_k}{\partial \bar{W}_{Lij}} \quad (39)$$

10

【数 3 8】

$$\left\{ \begin{array}{l} \frac{\partial a_L^n{}_k}{\partial \bar{W}_{Lij}} = \delta_{ki} \bar{x}_{L-1}^n{}_j \quad (40) \\ \Delta_L^n{}_k \equiv \frac{\partial E^n}{\partial a_L^n{}_k} = (a_L^n - t^n)_k \quad (41) \end{array} \right.$$

20

【0058】

また、最終の 1 つ前の重み層 (第 L - 1 層) の微分は、 $l = L - 1$ とおくと、下式 (42) ~ (48) の計算で求められる。

【数 3 9】

$$\forall i = D_l + 1, D_l + 2, \dots, D_l + D_L, \quad \frac{\partial E_l^n}{\partial \bar{W}_{lij}} = \sum_{k=D_l+1}^{D_l+D_L} \Delta_l^n{}_k \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} \quad (42)$$

30

【数 4 0】

$$\forall k = D_l + 1, D_l + 2, \dots, D_l + D_L$$

$$\left\{ \begin{array}{l} \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} = \delta_{ki} \bar{x}_{l-1}^n{}_j \quad (43) \\ \Delta_l^n{}_k \equiv \frac{\partial E_l^n}{\partial a_l^n{}_k} = h'(a_l^n)_k (h(a_l^n) - t^n)_k \quad (44) \end{array} \right.$$

40

【数 4 1】

$$\forall i = 1, 2, \dots, D_l, \quad \frac{\partial E_{l+1}^n}{\partial \bar{W}_{lij}} = \sum_{k=1}^{D_l} \Delta_l^n{}_k \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} \quad (45)$$

【数 4 2】

$$\forall k = 1, 2, \dots, D_l \quad 10$$

$$\left\{ \begin{array}{l} \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} = \delta_{ki} \bar{x}_{l-1}^n{}_j \quad (46) \\ \Delta_l^n{}_k \equiv \frac{\partial E_{l+1}^n}{\partial a_l^n{}_k} = \sum_{j=1}^{D_{l+1}} \frac{\partial E_{l+1}^n}{\partial a_{l+1}^n{}_j} \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} = \sum_{j=1}^{D_{l+1}} \Delta_{l+1}^n{}_j \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} \quad (47) \\ \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} = W_{l+1}{}_{jk} h'(a_l^n{}_k) \quad (48) \end{array} \right. \quad 20$$

ここで、式(41)はPNに直結する重みについての微分であり、式(45)はPNに直結しない重みについての微分である。

【0059】

最終の2つ前の重み層(第L-2層)の微分は、 $l = L - 2$ とおくと、下式(49)~(55)の計算で求められる。

【数 4 3】

$$\forall i = D_l + 1, D_l + 2, \dots, D_l + D_L, \quad \frac{\partial E_l^n}{\partial \bar{W}_{lij}} = \sum_{k=D_l+1}^{D_l+D_L} \Delta_l^n{}_k \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} \quad (49) \quad 30$$

【数 4 4】

$$\forall k = D_l + 1, D_l + 2, \dots, D_l + D_L$$

$$\left\{ \begin{array}{l} \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} = \delta_{ki} \bar{x}_{l-1}^n{}_j \quad (50) \\ \Delta_l^n{}_k \equiv \frac{\partial E_l^n}{\partial a_l^n{}_k} = h'(a_l^n{}_k) (h(a_l^n{}_k) - t^n)_k \quad (51) \end{array} \right. \quad 40$$

【数 4 5】

$$\forall i = 1, 2, \dots, D_l, \quad \frac{\partial E_{l+2}^n}{\partial \bar{W}_{lij}} + \frac{\partial E_{l+1}^n}{\partial \bar{W}_{lij}} = \sum_{k=1}^{D_l} \Delta_l^n{}_k \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} \quad (52)$$

【数 4 6】

$$\forall k = 1, 2, \dots, D_l \quad 10$$

$$\left\{ \begin{array}{l} \frac{\partial a_l^n{}_k}{\partial \bar{W}_{lij}} = \delta_{ki} \bar{x}_{l-1}^n{}_j \quad (53) \\ \Delta_l^n{}_k \equiv \frac{\partial E_{l+2}^n}{\partial a_l^n{}_k} + \frac{\partial E_{l+1}^n}{\partial a_l^n{}_k} \\ = \sum_{j=1}^{D_{l+1}} \frac{\partial E_{l+2}^n}{\partial a_{l+1}^n{}_j} \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} + \sum_{j=D_{l+1}+1}^{D_{l+1}+D_L} \frac{\partial E_{l+1}^n}{\partial a_{l+1}^n{}_j} \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} \quad 20 \\ = \sum_{j=1}^{D_{l+1}+D_L} \Delta_{l+1}^n{}_j \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} \quad (54) \\ \frac{\partial a_{l+1}^n{}_j}{\partial a_l^n{}_k} = W_{l+1}{}_{jk} h'(a_l^n{}_k) \quad (55) \quad 30 \end{array} \right.$$

【0060】

以上の導出より、 $l = L - 1, L - 2, \dots, 2$ においては、以下の通り更新則を一般化できる。

【数 4 7】

$$\forall i = 1, 2, \dots, D_l + D_L, \quad \frac{\partial E}{\partial \bar{W}_{lij}} = \sum_{n=1}^N \Delta_l^n{}_i \bar{x}_{l-1}^n{}_j \quad (56) \quad 40$$

【数 4 8】

$$\Delta_l^n{}_k = \begin{cases} \sum_{j=1}^{D_{l+1}+D_L} \Delta_{l+1}^n{}_j W_{l+1}{}_{jk} h'(a_l^n{}_k), & \forall k = 1, 2, \dots, D_l \\ h'(a_l^n{}_k) (h(a_l^n{}_k) - t^n)_k, & \forall k = D_l + 1, D_l + 2, \dots, D_l + D_L \end{cases} \quad (57)$$

そして、 $l = 1$ のときは、以下のとおりとすればよい。

【数 4 9】

$$\Delta_1^n_k = \sum_{j=1}^{D_2+D_L} \Delta_2^n_j W_{2jk} h'(a_1^n)_k \quad (58)$$

【0061】

次に、層数調整部 1 3 の処理について説明する。上記のようにして学習部 1 2 において、学習によってコスト E を最小化する

【数 5 0】

$$\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L$$

が求まると、層数調整部 1 3 は、そのコスト E に基づいて、ネットワークの上層を除去する。

【0062】

図 5 は、層数と学習終了時のテストコストとの関係を示す図である。図 5 におけるテストコストとは、テストデータセット 1 1 2 を用いて式 (3 3) を計算した場合の P N のコスト E のことであり、この値は小さい方が望ましい。図 5 の例では、簡単のため、出力層のユニット数及び各隠れ層の P N は 1 つとしている。

【0063】

学習部 1 2 における学習にて得られたネットワークの P N は、そのそれぞれがターゲット値に近づくように学習されているため、出力層のユニットと同じように、識別器と考えることができる。学習部 1 2 にて初期ネットワークを学習することで、図 5 のように各 P N のコストの値を得ることができる。テストデータに対して層数が多すぎれば P N は過学習を起し、層数が少なすぎれば P N は十分な弁別能力を持たなくなる。よって、P N のコストが最小値を取る層が最良の層である。図 5 の例では、第 7 層において、P N のコストが最小値をとっている。

【0064】

層数調整部 1 3 は、重み層の数をカウントする。P N コストが最小（最良）である層の後段の層は、全体の識別性能を向上させないため、計算不要である。従って、層数調整部 1 3 は、図 5 の点線で囲まれた部分をネットワークから除去する。即ち、層数調整部 1 3 は、コストが最小である P N より上層の P N は除去し、除去される P N に連結している重み（P N の前段の全ての重み）も除去する。

【0065】

図 6 は、層を除去した後のネットワークを示す図である。図 6 に示すように、第 7 層の P N が出力層のユニットとされる。なお、不要な上層をネットワークから除去することは、残ったネットワークの信号を一切変化させないため、P N のコストもまた変化しない。このことは、層数調整部 1 3 による不要な上層の除去が、識別性能を一切劣化させないことを意味している。

【0066】

次に、再学習部 1 4 における処理について説明する。再学習部 1 4 は、層数調整部にて上層を除去されたネットワークについて再度上記で説明した学習（ファインチューニング）を行う。図 7 は、再学習部 1 4 による処理を説明する図である。層数調整部 1 3 にて不要な層が除去されたネットワークは、さらなる学習によってコストを低減させることが可能である。ここでの学習は、現在保持している重みの値から、誤差逆伝搬法による更新を行うことを意味している。不要な層が除去されたネットワークは、もはや上層からの誤差の逆伝搬がないため、更新量は変化し、コストをさらに低減させることができる。

【0067】

再学習の結果得られるネットワークとその重みは、初期ネットワークの学習結果と比べ

10

20

30

40

50

て高い汎化能力を持つことが多い。その理由は、各隠れ層にPNを配置し、各隠れ層の線形分離性能を最適化することで、隠れ層の信号がデータの弁別のために必要以上に複雑な表現を持つことを回避し、訓練データへの過剰な適合を防いでいるためと考えられる。

【0068】

再学習部14は、再学習を行って、そのコストを計算する。そうすると、コストは低減するが、このとき最上層（出力層）のコストが最低になるとは限らない。例えば、再学習の結果、最上層の1つ前の層でコストが最低となることもあり得る。そこで、層数調整部13は、再学習後に最上層にてコストが最低となっていない場合には、コストが最低となっている層より上層の層を除去する。

【0069】

再学習部14は、層数調整部13にて上層が除去された場合には、さらにそのようなネットワークについて、再学習を行う。このような層数調整部13による不要な層の除去と再学習部14による再学習は、不要な層が出現しなくなるまで繰り返される。不要な層が出現しなくなると、PDNN学習装置100は、そのPDNNを最終的な結果として出力する。

【0070】

図8は、本実施の形態のPDNN学習装置によるPDNN学習方法のフロー図である。まず、学習部12は、初期ネットワークについてコスト計算による学習を行う（ステップS81）。次に、層数調整部13は、コストが最小である層より上層の層を除去して、残った最上層のPN、即ちコストが最小となるPNを出力層とする（ステップS82）。次に、再学習部14は、残った層からなるネットワークについて再学習を行う（ステップS83）。

【0071】

層最適化部13は、再学習を行った結果、依然として最上層のコストが最小であるかを判断する（ステップS84）。最上層のコストが最小でない場合には（ステップS84にてNO）、コストが最小である層より上層の層を除去し（ステップS82）、再学習部14はそのようなネットワークについて再学習を行う（ステップS83）。ステップS82～S84を繰り返して、ステップS84にて最上層のコストが最小となったときは（ステップS84にてYES）、その結果を出力する（ステップS85）。

【0072】

以上のように、本実施の形態のPDNN学習装置100によれば、層数調整部13が学習部12における学習に基づいて、PDNNの不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できる。

【0073】

また、本実施の形態のPDNN装置100によれば、PDNNの不要な上層を除去した後、再学習によってコストを低減するための学習を行うので、コストをより低減できる。さらに、再学習の後に再度上層の除去を試み、可能な場合に上層を除去し、さらに再学習を行い、これを上層の除去が有効である限り繰り返すので、さらにコストを低減して汎化能力の高いネットワークを構成できる。

【0074】

以下、上記の実施の形態に対する変形例を説明する。

【0075】

（変形例1）

上記の実施の形態では、層数調整部13は、学習部12にて学習の結果計算されたコストが最小になる層数を残して上層のユニットを除去した。しかしながら、再学習部14による再学習の結果、残った層のうちの最上層（図5の第7層）よりも上層の層でコストが最小になる可能性がある。そこで、本変形例のPDNN学習装置では、層数調整部13は、最上層（出力層）のコストが最小でない場合に、学習部12にて学習の結果計算されたコストが最小となる層より上のすべての層を直ちに除去するのではなく、コストが最小で

10

20

30

40

50

あるPNを有する層よりも上層の少なくとも1つのPNを残して上層を除去する。具体的には、層数調整部13は、最上層のコストが最小でない場合に、最上層から1層ずつ除去していく。

【0076】

層数調整部13が最上層を1層除去すると、再学習部14は、最上層が除去されたネットワークについて再学習を行い、そのコストを計算する。その結果、最上層のコストが最小となっていない場合には、層数調整部13は、さらに最上層を1層除去し、再学習部14は、そのようなネットワークについて再学習を行う。このようにして、最上層のコストが最小となるまで、最上層の除去及び再学習が繰り返される。

【0077】

上記のように、本変形例では、層数調整部13は、学習及びコスト計算によって不要と判断された層を直ちにすべて除去するのではなく、最上層から1層ずつ除去し、層が1層除去されるごとに、再学習部14にて再学習を行う。

【0078】

図9は、変形例のPDNN学習装置におけるPDNN学習方法のフロー図である。まず、学習部12は、上記の実施の形態と同様に、初期ネットワークについてコスト計算による学習を行う(ステップS91)。次に、層数調整部13は、コスト計算の結果、最上層においてコストが最小となっているか否かを判断し(ステップS92)、最上層においてコストが最小となっていない場合には(ステップS92にてNO)、最上層を除去する(ステップS93)。最上層が除去されると、再学習部14は、そのようなネットワークについて再学習を行い、コストを計算する(ステップS94)。

【0079】

層数調整部13は、再学習部14で再学習が行われた結果について、再び、最上層においてコストが最小となっているか否かを判断し(ステップS92)、最上層においてコストが最小となるまでステップS92～S94を繰り返す。ステップS92において、最上層においてコストが最小となった場合には(ステップS92にてYES)、そのネットワークを出力する(ステップS95)。

【0080】

即ち、層数調整部13は、最上層のPNのコスト(図5の例の場合は第9層のコスト)が最小でない場合に、コストが最小であるPNを有する層(図5の例の場合は第7層)よりも上層の少なくとも1つのPN(図5の例の場合は第8層のPN)を残して上層(図5の例の場合は、第8層のPN以外のユニット及び第9層)を除去する。このように、本変形例によれば、層を1層ずつ除去していくので、複数回の再学習によって、最適な層数を決定することができる。

【0081】

(変形例2)

次に、変形例2として、出力層に順伝搬しない部分が層構造を持つ場合を説明する。図10は、本変形例のネットワーク構造を示す図である。図10に示すように、各層のユニット数は3であるが、この3ユニットから直ちにPNを得るのではなく、PNを得る前に、さらに複数のニューロンを経て、それらのニューロンからPNを算出する。

【0082】

本変形例においても、上記の実施の形態と同様に、PNの識別性能によって層数を減らすことが可能である。また、PNに接続する部分が層構造をなすことで、PNの識別性能を高めることができると考えられる。さらに、本変形例では、各隠れ層の線形分離性能の最適化の制約がないので、出力層における識別性能の向上も期待できる。

【0083】

以下、本実施の形態及びその変形例のPDNN学習装置の応用例を説明する。

【0084】

(応用例1)

図11は、応用例1の多層ニューラルネットワーク学習装置の構成を示すブロック図で

10

20

30

40

50

ある。多層ニューラルネットワーク学習装置 200 は、上記の実施の形態の PDNN 学習装置 100 と、PN のない通常の高層ニューラルネットワークの学習部 101 を備えている。学習部 101 は、従来と同様の学習方法によって多層ニューラルネットワークの学習を行う。PDNN 学習装置 100 は、この学習部 101 に対して、PDNN の学習の結果（除去の後に残った重みの値）を初期値として与える。

【0085】

PDNN は、各隠れ層の線形分離性能が最適化されているため、過度に非線形な内部表現を生成しない利点がある。一方、PN のないネットワークと比較して、PN のあるネットワークでは、各ユニットの役割が大きく異なる。即ち、PN のないネットワークにおいて、あるユニットは、出力層において最高の性能となるように信号を調整するのに対し、PN のあるネットワークにおいて、あるユニットは、それ以降すべての層において良い識別性能を与えるように調整される。この強い制約により、PDNN の出力層の識別性能は制限を蒙っている可能性がある。

10

【0086】

PDNN 学習装置 100 は、上記の実施の形態のとおり、PDNN の学習、不要な層の除去、及び再学習を行い、不要な層が出現しなくなるまで上層の除去と再学習を繰り返す。この時の重みを保持したまま、出力ユニットへ順伝搬しないすべてのユニットと重み（図 1 においては PN と、PN に連結する重み）をネットワークから除去して、このネットワークを学習部 101 の初期ネットワークとして出力する。学習部 101 は、通常の高層ニューラルネットワークの学習を行う。

20

【0087】

（応用例 2）

本応用例では、上記の実施の形態又はその変形例によって学習された PDNN を用いた識別器について説明する。適切に学習された PDNN によって早期棄却を使った高速な識別が可能である。本応用例の識別器は、PDNN を利用して早期棄却を行う。本応用例では、早期棄却の考え方を画像による歩行者検知を例にとりて説明する。

【0088】

図 13 は、車載カメラを使って歩行者を検知する例を示す図である。識別器は、図 13 に示すように、画像に対し識別器のウィンドウをスキャンすることで領域を網羅的に識別する（ターゲットであるかないかを判定する）。図 13 の例のように、ウィンドウによって囲まれる領域の大半が背景である場合、例えば空や路面といった明らかに歩行者ではない部分の識別が高速化されれば、画像全体の識別が非常に高速化されることになる。このようにターゲットではないエリアにおける判定を少ない演算量で棄却することを早期棄却という。また、ターゲットがそうでないかの判定が難しい領域は、より演算量を費やした方法により、より信頼性の高い識別を行う。このことにより、全体として高い識別性能を保持することが可能である。

30

【0089】

図 14 は、PN を利用した早期棄却の処理を説明する図である。適切に学習された PDNN は、各 PN が識別器の役割を持つため、早期棄却が可能である。最初の PN の値が計算された時点で、背景であるスコアが十分高ければ（PN の出力がある閾値以下又はある閾値以上であれば）、その時点で識別結果を「背景」とし、その層より上層の計算を棄却する。このルールを全ての隠れ層の PN に適用することで、段階的な棄却が可能となる。

40

【0090】

具体的には、識別器は、各隠れ層の PN を算出するたびに、それが閾値以下となっているか否かを判断し、閾値以下となっている場合は、直ちに歩行者なしと識別する。閾値以下となっていない場合には、次の層の計算を行う。ウィンドウの大半にターゲットが存在しない場合には、どれほど層の多い PDNN であっても、早期棄却が行われることにより、実際の計算時間はそれよりはるかに少ない層を持った多層ニューラルネットワーク相当となる。

【0091】

50

(応用例 3)

応用例 3 は、応用例 2 の変形例である。本応用例では、出力層のユニット及び各層の P N に異なる種類のターゲット値を近似させることにより、早期棄却を行う。図 1 5 は、本応用例の早期棄却の処理を示す図である。以下、例として、車載カメラ画像からの標識認識の問題に対し、図 1 5 の P D N N の適用を考える。

【0092】

P N に対してはある画像パッチの中に標識があるか・ないかの二値分類を学習させることとする。標識はいかなる種類でもよいこととする。一方、出力層においては、止まれの標識があるか・ないかの二値分類を行うユニット、駐車禁止の標識があるか・ないかの二値分類を行うユニット、といった具合に、それぞれ固有の標識を認識できるように学習させたユニットを配置する。

10

【0093】

このネットワークを使えば、低層においては、標識のある・なしの二値分類のみを行うこととなり、全ての種類の標識の認識を網羅的に行うよりはるかに低い演算量で棄却が可能である。実際に画像パッチに標識がある場合においては、出力層において網羅的な識別を行うことができるため、種類の認識を的確に行うことができる。なお、各 P N における早期棄却のための閾値は、画像全体の計算時間と識別性能のバランスを取りながらチューニングして決定する。

【0094】

なお、本応用例の識別器のために P D N N 学習装置 1 0 0 にて学習を行う場合には、各隠れ層は、とまれ、駐車禁止、徐行等の出力層のすべてに対応した P N を有することになる。これらの P N のすべてを識別器においても早期棄却のためのニューロンとして利用してよい。こうすることで、隠れ層の各層において、とまれ、駐車禁止、徐行、...等の標識の有無を判断して早期棄却を行うことができる。

20

【0095】

さらに、隠れ層の下層で早期棄却ができる可能性が低い場合には、隠れ層のうちの途中の層から早期棄却のためのニューロンを用意してもよい。これによって、早期棄却のための効率の低い演算に演算資源および時間を割くことを回避することができる。

【0096】

(応用例 4)

本応用例では、P D N N 学習装置 1 0 0 で学習された P D N N を用いて手書き数字を識別する識別器を説明する。図 1 6 は、手書き数字のデータセットを示す。各数字は、2 8 × 2 8 画素のグレイスケールであり、正解は 0 ~ 9 までの 1 0 種類ある。図 1 7 は、P D N N のネットワーク構造を示す図であり、図 1 8 は、P N なしの従来の学習によって得られた多層ニューラルネットワークの構造を示す図である。

30

【0097】

本応用例の識別器で用いる P D N N の各層のユニット数は 1 0 0 であり、出力層は上述のように 1 0 種類である。隠れ層には 1 0 個の P N が用意されており、各層において、早期棄却の判定が行われる。入力データとしては、2 8 × 2 8 の画素の白黒に対応するバイナリコードが入力される。

40

【0098】

図 1 9 は、P N を用いて学習をしたネットワーク (「with probing neuron (no fine tuning)」、P N を用いて学習をし、さらにファインチューニングを行ったネットワーク (「with probing neuron (fine tuning)」、及び図 1 8 に示すように P N なしで学習を行ったネットワーク (「without probing neuron」) のそれぞれについて、図 1 6 のデータセットを用いてテストをした結果を示すグラフである。図 1 9 は、層数に対するコストを示している。ここで、コストとしては、クロスエントロピーを用いている。図 1 9 に示すように、第 3 層以降は、P N なしの場合と比較して、P N ありの学習のほうがコストが低く、ファインチューニングを行った場合にはさらにコストが低減されている。

【0099】

50

(応用例 6)

本応用例では、上記の実施の形態の P D N N 学習装置 1 0 0 を畳み込みニューラルネットワーク (C N N : Convolutional Neural Network) の学習装置として応用する例を説明する。

【 0 1 0 0 】

まず、C N N について説明する。C N N については、例えば、Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network", Advances in Neural Information Processing Systems (NIPS), pp. 396-404, 1990. (非特許文献 2) に C N N を使った画像認識手法が記載されている。

10

【 0 1 0 1 】

C N N の学習において使用するデータについて説明する。教師データとして、ネットワークの入力となる 2 次元配列の画像と、それに対応した、ネットワークの出力となるターゲット値 (多次元ベクトルまたはスカラー) の組が多数与えられているものとする。入力となる 2 次元配列の画像は、グレイスケールであってもよいし、また R G B などの要素を持つカラー画像であってもよい。カラー画像の場合であれば、入力は 3 つの 2 次元配列となる。上記の実施の形態と同様に、教師データは、ネットワークを学習するために使用される訓練データセットと、学習されたネットワークを評価するテストデータセットに分割しておくものとする。

【 0 1 0 2 】

図 2 0 は、C N N の典型的なネットワーク構造を示す図である。まず、入力画像に対しフィルタによる畳み込み (convolution) と、画像の解像度を下げる操作であるプーリング (pooling) が施される。この畳み込みとプーリングを数回繰り返し、十分に解像度が小さくなった段階で、各ピクセルを入力ユニットとする順伝搬型の多層ニューラルネットワークに結合される。なお、図 2 0 では隠れ層の図示を省略している。

20

【 0 1 0 3 】

図 2 0 中の網掛けされた要素はフィルタであり、フィルタとは n (ピクセル) \times n (ピクセル) の要素を持つ重みである。なお、このフィルタには、バイアスを加えてもよい。この C N N を誤差逆伝搬法により学習することで、局所的な歪みに頑健でありかつ識別に有効な特徴を抽出できるようになる。畳み込み演算された値は、シグモイド関数などの活性化関数 (activation function) によって、非線形写像が行われる (図示は省略) 。こうして得られた信号は、入力と同様に 2 次元配列で記述できる。

30

【 0 1 0 4 】

次の演算は、プーリング (pooling) である。典型的なプーリングは、上述の二次元配列を 2×2 のグリッドに分割し、各グリッドの 4 つの信号の最大値を取る処理である。この処理により、上述の二次元配列はその 4 分の 1 のサイズに縮小される。プーリングにより画像の位置に関する特徴を失わずに情報を圧縮することが可能となる。プーリングの結果得られた 2 次元配列をマップ (map) と呼ぶ。マップの集まりが畳み込みニューラルネットワークにおいて 1 つの隠れ層を形成する。

【 0 1 0 5 】

上記の例とは異なるプーリングも可能である。例えば、プーリングとして、 2×2 のグリッドから最大値を抽出する最大値プーリング (max pooling) 、グリッドのある 1 つの要素 (例えば (1 , 1) 要素) だけを取り出すサブサンプリング (subsampling) 等を採用してもよい。また、グリッドをオーバーラップするようにしてもよい。いずれにしても、畳み込みされた 2 次元配列を縮小させる点は同じである。

40

【 0 1 0 6 】

なお、畳み込みとプーリング (convolution + pooling) は、通常複数回反復する。図 2 0 の例では二重であるがそれ以上の繰り返し回数であってもよい。畳み込みとプーリングを反復させて情報を十分に圧縮した後、(畳み込みではない) 通常のニューラルネットワークを適用する。

50

【 0 1 0 7 】

図 2 0 の例では、畳み込みとプーリングを 2 度行った結果得られる 2 次元配列の各要素を 1 つのユニットとして、通常のニューラルネットワークが連結されている。この通常のニューラルネットワークは多層であってよい。

【 0 1 0 8 】

以上のように、CNN は順伝搬型の多層ニューラルネットワークの一種である。即ち、ある層の信号は、その 1 つ前の層の信号と両者の間の重みの関数であり、また、この関数は微分が可能である。従って、CNN には誤差逆伝搬法が適用可能である。そこで、本応用例では、上記の実施の形態と同様に CNN に PN を導入する。以下、PN が導入された CNN を、プローブ畳み込みニューラルネットワーク (PCNN: Probed Convolutional Neural Network) といい、その学習装置を PCNN 学習装置という。PCNN は、ターゲット値に近づくように学習されたユニットを各隠れ層に持つ CNN である。

10

【 0 1 0 9 】

図 2 1 は、PCNN のネットワーク構造を示す図である。図 2 1 に示すように、PCNN では、マップに対して PN を連結する。マップから連結される PN は、マップのすべての画素と連結していてもよいし、縦横 1 画素おきの結合を有していてもよい。

【 0 1 1 0 】

PCNN 学習装置は、上記の実施の形態の PDNN 学習装置 1 0 0 と同様に、教師データ記憶部 1 1、学習部 1 2、層数調整部 1 3、及び再学習部 1 4 を備え、拡張された誤差逆伝搬法により学習を行う。さらに、PCNN 学習装置の層数調整部 1 3 は、PDNN 学習装置 1 0 0 と同様に、全結合された層のうち、不要な上層を特定してネットワークから除去する。PCNN 学習装置は、複数の畳み込み (+ プーリング) 層において、第 n 層の PN の識別性能が第 $n - 1$ 層の PN の識別性能を上回らなかった場合は、第 n 層の畳み込み (+ プーリング) 層を除去したネットワークを生成し、第 n 層以降の重みを乱数によって初期化して PCNN の学習を繰り返してもよい。

20

【 0 1 1 1 】

また、PCNN 学習装置の再学習部 1 4 は、PDNN 学習装置 1 0 0 と同様に、除去されたネットワークを再学習する。これによってコストをさらに低減させることができる。また、PN を持たない CNN の良い初期値を与える方法として、PCNN 学習装置で学習された PCNN を利用することも可能である。さらに、PCNN も上記の応用例と同様に、識別器において早期棄却による高速な検出が可能である。

30

【 産業上の利用可能性 】

【 0 1 1 2 】

以上のように、本発明は、層数調整部が学習部における学習に基づいて、不要な上層を除去するので、層数を無駄に多くして処理負荷及び計算に要する時間を増大させることなく、十分な汎化能力を有する多層ニューラルネットワークを取得できるという効果を有し、教師あり学習によって順伝搬型の多層ニューラルネットワークの学習を行う学習装置等として有用である。

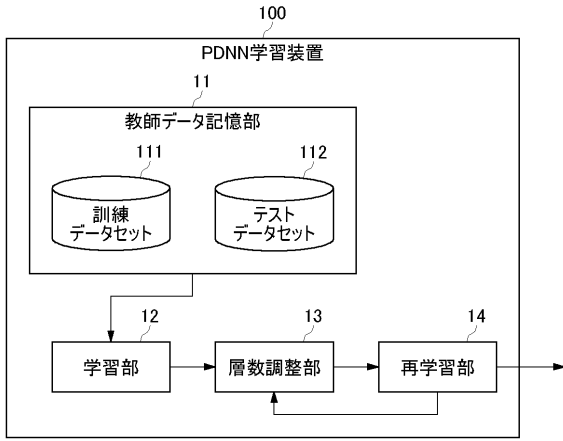
【 符号の説明 】

【 0 1 1 3 】

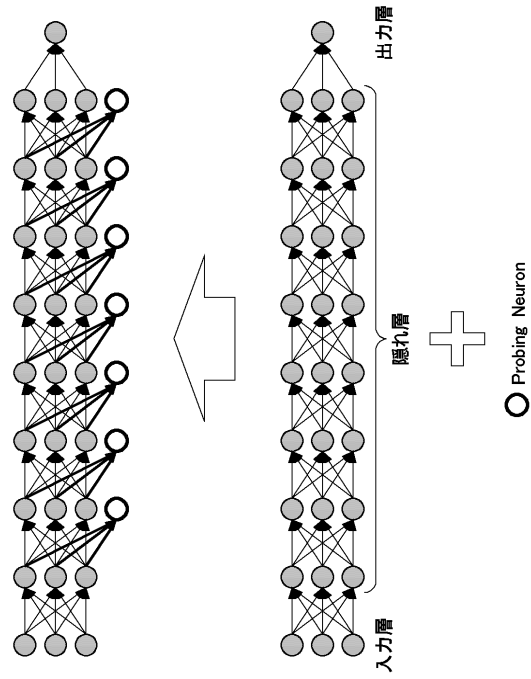
- 1 0 0 PDNN 学習装置 (多層ニューラルネットワーク学習装置)
- 1 1 教師データ記憶部
- 1 2 学習部
- 1 3 層数調整部
- 1 4 再学習部
- 1 0 1 学習部
- 2 0 0 多層ニューラルネットワーク学習装置

40

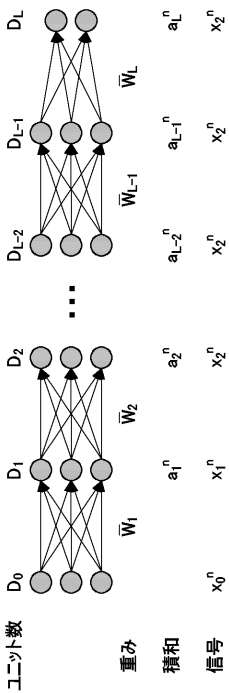
【 図 1 】



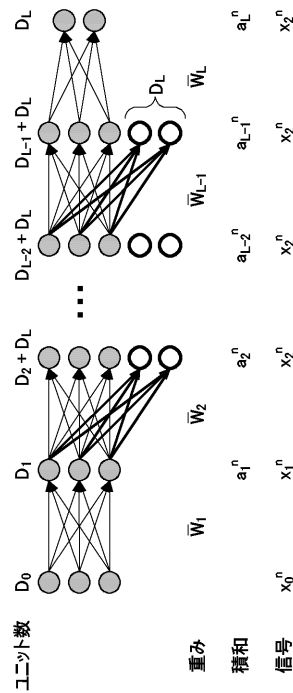
【 図 2 】



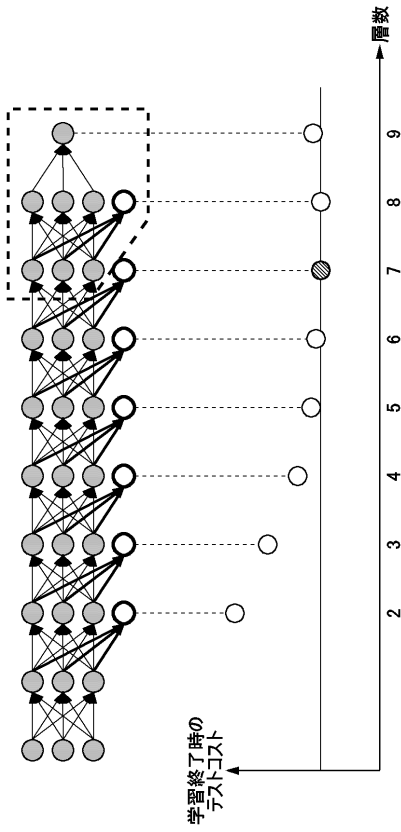
【 図 3 】



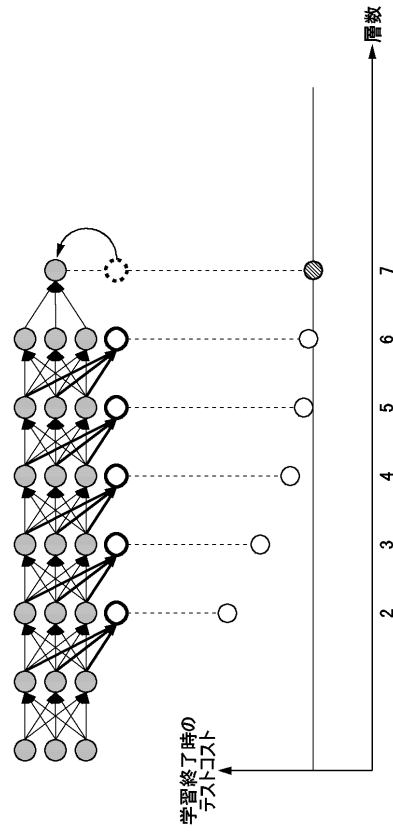
【 図 4 】



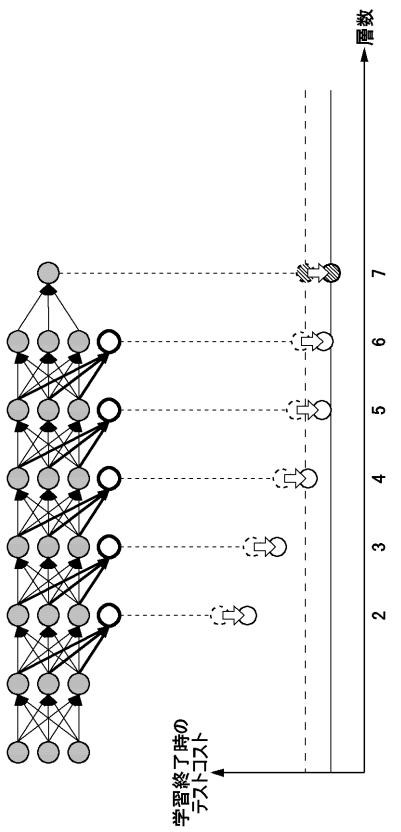
【図5】



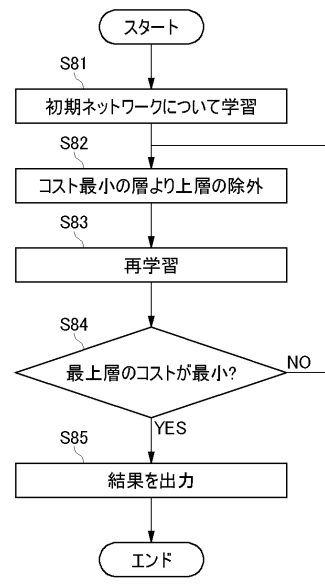
【図6】



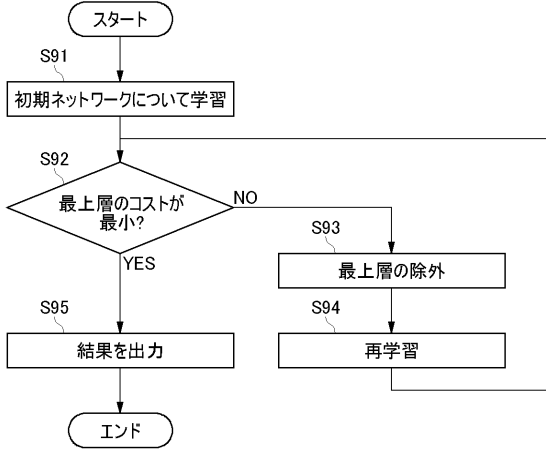
【図7】



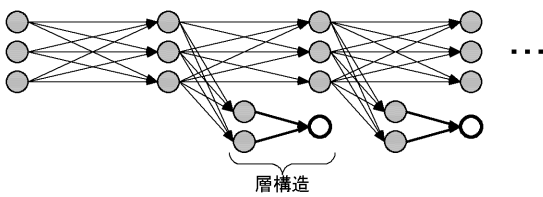
【図8】



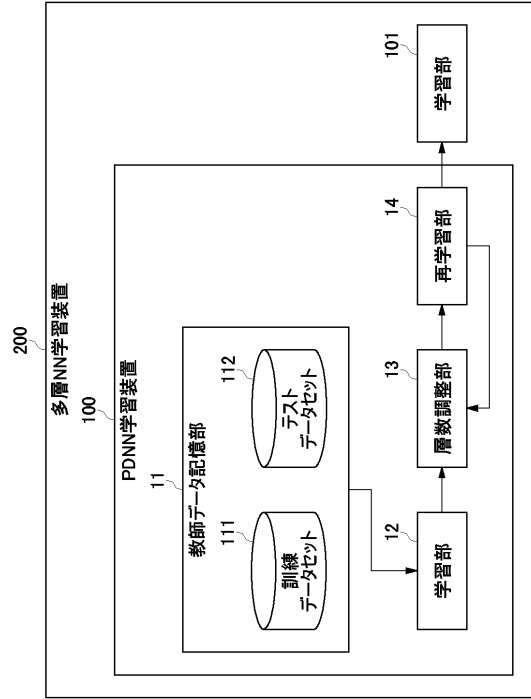
【図9】



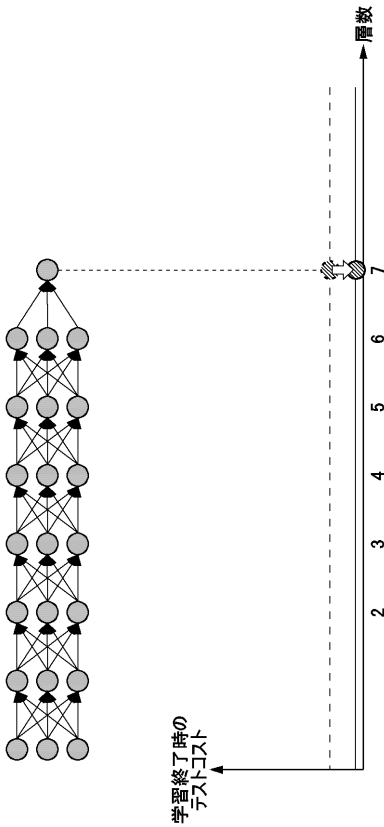
【図10】



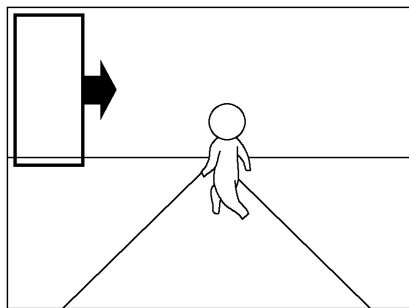
【図11】



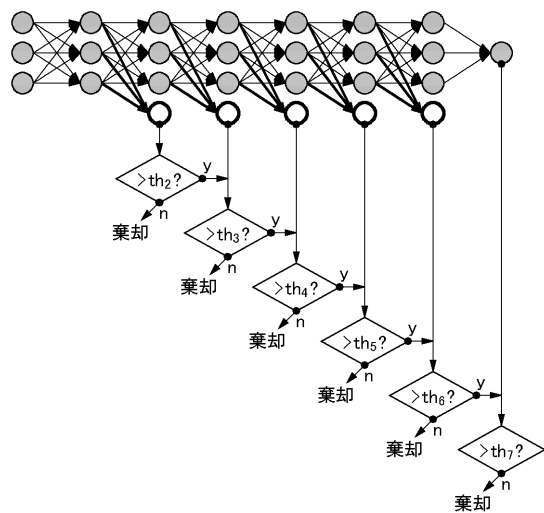
【図12】



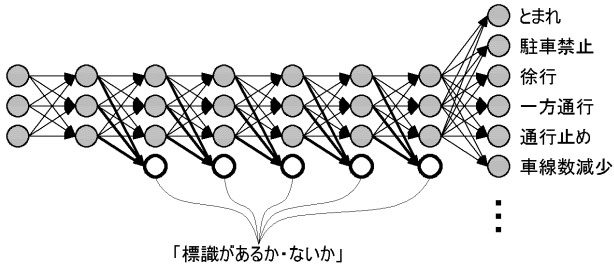
【図13】



【図14】



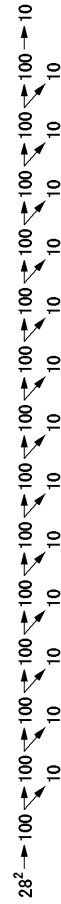
【 図 1 5 】



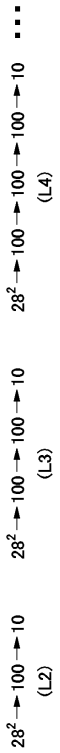
【 図 1 6 】



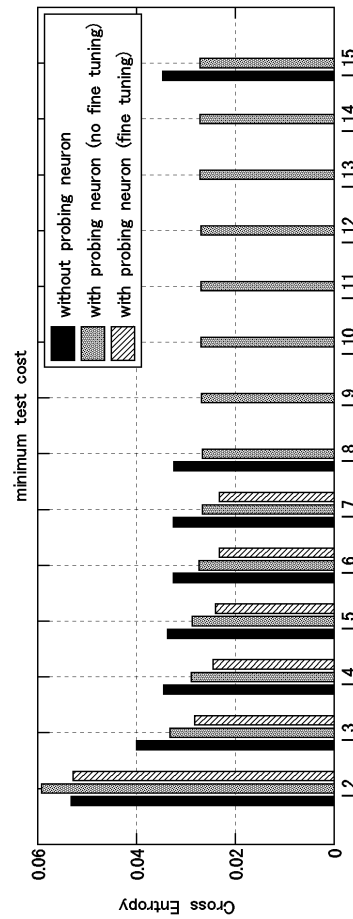
【 図 1 7 】



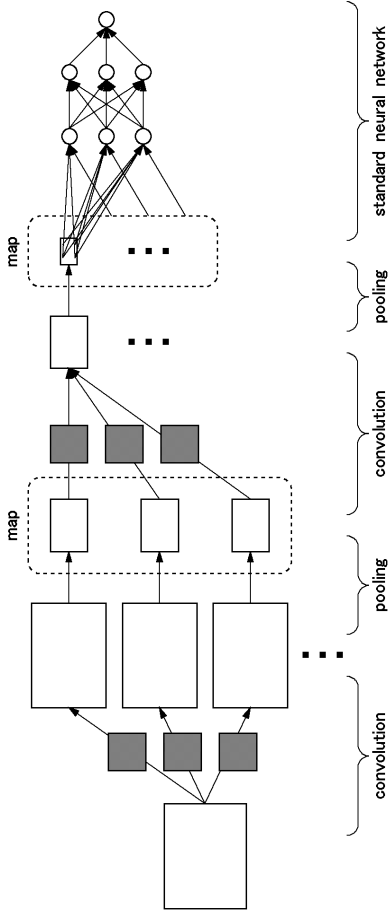
【 図 1 8 】



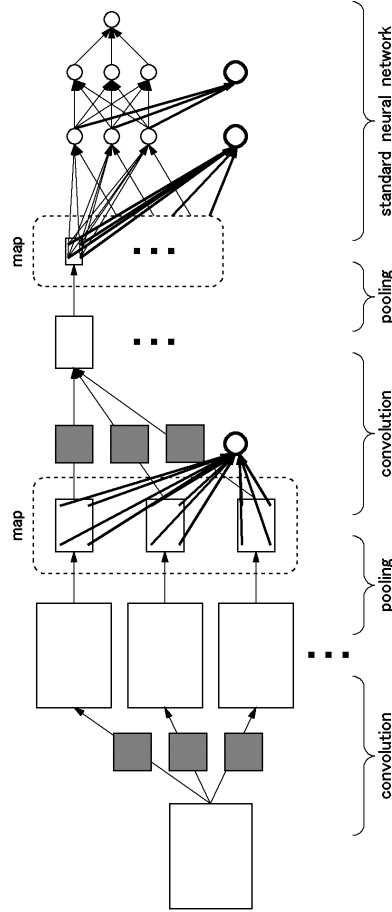
【 図 1 9 】



【 図 2 0 】



【 図 2 1 】



フロントページの続き

(74)代理人 100115808

弁理士 加藤 真司

(72)発明者 佐藤 育郎

東京都渋谷区渋谷二丁目15番1号 渋谷クロスタワー28F 株式会社デンソーアイティールボ
ラトリ内

(72)発明者 玉津 幸政

愛知県刈谷市昭和町一丁目1番地 株式会社デンソー内