

(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
2017年8月31日(31.08.2017)



(10) 国際公開番号

WO 2017/145852 A1

- (51) 国際特許分類:
G06N 3/08 (2006.01)
- (21) 国際出願番号: PCT/JP2017/005230
- (22) 国際出願日: 2017年2月14日(14.02.2017)
- (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 2016-032806 2016年2月24日(24.02.2016) JP
- (71) 出願人: 日本電気株式会社(NEC CORPORATION)
[JP/JP]; 〒1088001 東京都港区芝五丁目7番1号
Tokyo (JP).
- (72) 発明者: 石井 雅人(ISHII Masato); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内
Tokyo (JP).
- (74) 代理人: 下坂 直樹(SHIMOSAKA Naoki); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内
Tokyo (JP).
- (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, JP, 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.

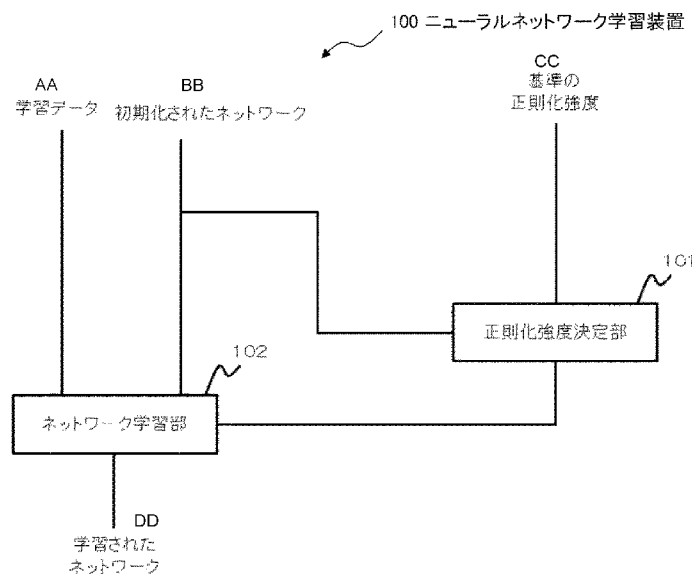
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーロアジア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ユーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, 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:

— 国際調査報告 (条約第 21 条(3))

(54) Title: NEURAL NETWORK LEARNING DEVICE, NEURAL NETWORK LEARNING METHOD AND STORAGE MEDIUM STORING PROGRAM

(54) 発明の名称: ニューラルネットワーク学習装置、ニューラルネットワーク学習方法及びプログラムを記憶する記憶媒体



- 100 Neural network learning device
101 Normalized intensity determination unit
102 Network learning unit
AA Learning data
BB Initialized network
CC Reference normalized intensity
DD Learned network

(57) Abstract: With the present invention, an entire network can be efficiently learned. A neural network learning device of one embodiment of the present invention is characterized by being provided with a normalized intensity determination unit that determines a normalized intensity for each layer, on the basis of an initialized network, and a network learning unit that performs network learning on the basis of the initialized network and the normalized intensities determined by the normalized intensity determination unit, wherein the normalized intensity determination unit determines normalized intensities such that the difference between a parameter update amount size calculated from a loss function and a parameter update amount size calculated from a normalized term is within a prescribed range.

(57) 要約: ネットワーク全体を効率的に学習することができる。本発明の一態様におけるニューラルネットワーク学習装置は、初期化されたネットワークに基づいて層ごとの正規化強度を決定する正規化強度決定部と、初期化されたネットワークと前記正規化強度決定部で決定された正規化強度とに基づいてネットワークの学習を行うネットワーク学習部とを備え、前記正規化強度決定部は、損失関数から計算されるパラメータ更新量の大きさと正規化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正規化強度を決定することを特徴とする。

明 細 書

発明の名称：

ニューラルネットワーク学習装置、ニューラルネットワーク学習方法及びプログラムを記憶する記憶媒体

技術分野

[0001] 本発明は、ニューラルネットワーク学習装置、ニューラルネットワーク学習方法及びプログラムに関し、特に、深いネットワークでも効率よくネットワークを学習する装置、方法及びプログラムを記憶する記憶媒体に関する。

背景技術

[0002] パターン認識技術とは、入力されたパターンがどのクラスに属するパターンであるかを推定する技術である。具体的なパターン認識の例としては、画像を入力として写っている物体を推定する物体認識や、音声を入力として発話内容を推定する音声認識などが挙げられる。

[0003] パターン認識技術として統計的機械学習が広く利用されている。特にニューラルネットワークは、近年、`deep learning`と呼ばれる学習技術の発達により、深層ネットワークの学習が可能となり、入力パターンの変動に頑健な認識を行うことができることが知られている。

[0004] 深層ネットワークを学習する際に、一般的には正則化を行うが、適切な正則化強度を設定することが難しい。正則化とは、過学習を回避するため、学習されるパラメータが極端な値をとらないようにする処理である。

[0005] 非特許文献1は、ニューラルネットワークの学習に関する技術を開示する。非特許文献1が開示する技術では、例えば、パラメータの2乗和を正則化項とするL2正則化が用いられ、損失関数と正則化項の総和を減少させるように学習を行う。

先行技術文献

非特許文献

[0006] 非特許文献1: Yoshua Bengio, "Practical Re

commendations for Gradient-Based Training of Deep Architectures,” Neural Networks: Tricks of the Trade, 2nd Edition, Volume 7700 of the series Lecture Notes in Computer Science, pp 437–478, 2012.

発明の概要

発明が解決しようとする課題

[0007] 例えば、正則化の大きさが学習に与える影響について図3を用いて説明する。図3は正則化項の働きの一例を示す図である。簡単のため、2つのパラメータ w_1 , w_2 の場合について考える。正則化が弱すぎる場合、損失関数のみを低減する方向（図3のA）にパラメータが更新され、過学習が発生する。逆に、正則化が強すぎる場合、正則化項のみを低減する方向（図3のB）にパラメータが更新され、多くのパラメータがゼロに収束してしまい、学習が進まない。したがって、正則化の大きさを適切に設定することで更新方向を調整し（図3のC）、いずれの問題も起きないように更新を行う必要がある。

[0008] 非特許文献1が開示する技術では、各層で均一な強度の正則化を行うが、このような正則化では、正則化が強すぎる層と弱すぎる層が発生してしまう。なぜなら、例えば学習に勾配法を用いる場合、学習におけるネットワークの更新時、損失関数の勾配の大きさは誤差逆伝播により更新対象より上の層全てのスケールに依存するが、正則化項の勾配の大きさは更新対象の層自身のスケールにしか依存せず、両者の比は各層で不均一になってしまうためである。したがって、図4の左側の列で示すように、非特許文献1が開示する技術は、各層で均一な強度の正則化では、いずれかの層では正則化が弱すぎて過学習が発生したり、またいずれかの層では正則化が強すぎて学習が進まなかったりしてしまい、深層ネットワークを適切に学習することが難しかった。また、層ごとの正則化強度を個別に手動で設定する場合でも、深層ネッ

トワークでは層の数が多いため、試行すべき正則化強度の組み合わせが膨大になり、層ごとに最適な正則化強度を決定することは現実的ではなかった。

[0009] 本発明の目的の1つは、上記課題を解決し、ネットワーク全体を効率的に学習するニューラルネットワーク学習装置、ニューラルネットワーク学習方法及びプログラムを提供することである。

課題を解決するための手段

[0010] 本発明の一態様におけるニューラルネットワーク学習装置は、初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定手段と、初期化されたネットワークと前記正則化強度決定部で決定された正則化強度とに基づいてネットワークの学習を行うネットワーク学習手段とを備え、前記正則化強度決定手段は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定することを特徴とする。

[0011] 本発明の一態様におけるニューラルネットワーク学習方法は、初期化されたネットワークに基づいて層ごとの正則化強度を決定し、初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行い、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定することを特徴とする。

[0012] 本発明の一態様におけるプログラムは、コンピュータに、初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定処理と、初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行うネットワーク学習処理と、を実行させ、前期正則化強度決定処理は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定することを特徴とする。本発明は、上述のプログラムを記憶する記録媒体によっても実現できる。

発明の効果

[0013] 本発明によれば、ネットワーク全体を効率的に学習することができる。

図面の簡単な説明

[0014] [図1]本発明の実施形態に係るニューラルネットワーク学習装置の機能構成を示すブロック図である。

[図2]本発明の実施形態に係るニューラルネットワーク学習装置の動作を示すフローチャートである。

[図3]正則化項の働きの一例を示す図である。

[図4]本発明の実施形態におけるネットワークの各層で最適な強度の正則化の一例を示す図である。

[図5]本発明の実施形態に係るニューラルネットワーク学習装置を実現するコンピュータ装置のハードウェア構成を示すブロック図である。

発明を実施するための形態

[0015] 以下、図面を参照して、本発明の実施形態および実施形態の変形例を説明するが、本発明は本実施形態および本変形例に限定されるものではない。なお、以下で説明する図面で、同機能を有する要素には同一符号を付け、その繰り返しの説明は省略することもある。

[0016] 図5を用いて、本発明の実施形態に係るニューラルネットワーク学習装置100を構成するハードウェアについて説明する。図5は、本発明の実施形態に係るニューラルネットワーク学習装置100を実現するコンピュータ装置200のハードウェア構成を示すブロック図である。なお、本発明の実施形態において、ニューラルネットワーク学習装置100の各構成要素は、機能単位のブロックを示している。ニューラルネットワーク学習装置100の各構成要素は、例えば図5に示すようなコンピュータ装置200とソフトウェアとの任意の組み合わせにより実現することができる。

[0017] 図5が示すように、コンピュータ装置200は、CPU (Central Processing Unit) 201と、ROM (Read Only Memory) 202と、RAM (Random Access Memory) 203と、記憶装置205と、ドライブ装置207と、通信イン

ターフェース２０８と、入出力インターフェース２１０と、バス２１１とを備える。

[0018] 記憶装置２０５は、プログラム２０４を格納する。ドライブ装置２０７は、記憶媒体２０６の読み書きを行う。通信インターフェース２０８は、ネットワーク２０９と接続する。入出力インターフェース２１０は、データの入出力を行う。バス２１１は、各構成要素を接続する。

[0019] CPU２０１は、RAM２０３を用いてプログラム２０４を実行する。プログラム２０４は、ROM２０２に記憶されていてもよい。また、プログラム２０４は、記憶媒体２０６に記憶され、ドライブ装置２０７によって読み出されてもよいし、外部装置からネットワーク２０９を介して送信されてもよい。通信インターフェース２０８は、ネットワーク２０９を介して外部装置とデータをやり取りする。入出力インターフェース２１０は、周辺機器（キーボード、マウス、表示装置など）とデータをやり取りする。通信インターフェース２０８及び入出力インターフェース２１０は、データを取得又は出力する手段として機能することができる。出力情報などのデータは、記憶装置２０５に記憶されていてもよいし、プログラム２０４に含まれていてもよい。

[0020] なお、本発明の実施形態に係る各装置の実現方法には様々な変形例がある。例えば、本発明の実施形態に係る各装置は、専用の装置として実現することができる。また、本発明の実施形態に係る各装置は、互いに通信可能に接続された複数の装置の組み合わせにより実現することができる。

[0021] なお、以下の本発明の実施形態の機能を実現するように該実施形態の構成を動作させるプログラム（より具体的には、図２等を示す処理をコンピュータに実行させるプログラム）を記憶媒体に記憶させ、該記憶媒体に記憶されたプログラムをコードとして読み出し、コンピュータにおいて実行する処理方法も本発明の実施形態の範疇に含まれる。すなわち、コンピュータ読取可能な記憶媒体も実施形態の範囲に含まれる。また、上述のプログラムが記録された記憶媒体はもちろん、そのプログラム自体も本発明の実施形態に含ま

れる。

[0022] 該記憶媒体としては例えばフロッピー（登録商標）ディスク、ハードディスク、光ディスク、光磁気ディスク、CD（Compact Disc）－ROM、磁気テープ、不揮発性メモリカード、ROMを用いることができる。また該記憶媒体に記憶されているコード単体で処理を実行するプログラムに限らず、他のソフトウェア、拡張ボードの機能と共同して、OS（Operating System）上で動作して処理を実行するプログラムも実施形態の範疇に含まれる。

[0023] 次に、本発明の実施形態に係るニューラルネットワーク学習装置100の機能について説明する。図1は、本発明の実施形態に係るニューラルネットワーク学習装置100の機能構成を示すブロック図である。図1が示すブロックは単一の装置内に実装されてよく、あるいは複数の装置内に別れて実装されてよい。ブロック間のデータの授受は、データバス、ネットワーク、可搬記憶媒体等、任意の媒体を介して行われてよい。

[0024] 図1が示すように、本実施形態に係るニューラルネットワーク学習装置100は、初期化されたネットワークに基づいて各層の正則化強度を決定する正則化強度決定部101と、初期化されたネットワークを正則化強度決定部101で決定した正則化強度で学習するネットワーク学習部102とを含む。

[0025] 次に、本実施形態に係るニューラルネットワーク学習装置100の動作について説明する。図2は、本発明の実施形態に係るニューラルネットワーク学習装置100の動作例を示すフローチャートである。正則化強度決定部101は、初期化されたネットワークと基準の正則化強度とに基づいて、層ごとの正則化強度を決定し、決定した正則化強度をネットワーク学習部102に出力する（ステップS101）。ネットワーク学習部102は、学習データと初期ネットワークと正則化強度決定部101から入力される正則化強度とに基づいてネットワークの学習を行い、学習したネットワークを出力する（ステップS102）。

[0026] 以下、具体的な例を用いて本発明の実施形態に係るニューラルネットワーク学習装置 100 の動作を説明する。

[0027] 正則化強度決定部 101 は、初期化されたネットワークに基づき、層ごとの適切な正則化強度を算出する。正則化強度決定部 101 は、本例では、学習に勾配法を用いるとし、最終層の正則化強度を基準として、損失関数の勾配の分散と正則化項の勾配の分散との比が各層で等しくなるように正則化強度を決定する。ここで、 k 層目のパラメータ W_k に関する損失関数の勾配の分散を $\Delta_k E$ 、正則化項の勾配の分散を $\Delta_k R$ とすると、正則化強度決定部 101 は、以下を満たすように k 層目の正則化強度 λ_k を決定する。以下の式において、 L は最下層の番号である。 $\Delta_L E$ は、最下層（すなわち L 層目）のパラメータ W_L に関する損失関数の勾配の分散を表す。 $\Delta_L R$ は、 L 層目の正則化項の勾配の分散を表す。

[0028] [数1]

$$\frac{\lambda_k \cdot \Delta_k R}{\Delta_k E} = \frac{\lambda_L \cdot \Delta_L R}{\Delta_L E}$$

[0029] 例えば図 4 を用いて説明する。図 4 は、本発明の実施形態におけるネットワークの各層で最適な強度の正則化の一例を示す図である。図 4 に示す例では、数 1 の右辺が $5 / 5 = 1$ である。すべての λ_k が 1 であれば、 $k=1$ である場合の数 1 の左辺は、図 4 の左側に示すように、 $3 / 8$ である。さらに、 $k=2$ である場合の数 1 の左辺は、図 4 の左側に示すように、 $7 / 2$ である。この場合、各層で $\lambda_k \cdot \Delta_k R / \Delta_k E = 1$ となるように λ_k を決定すると、 $\lambda_1 = 3 / 8$ 、 $\lambda_2 = 7 / 2$ となる。これにより、損失関数の勾配の大きさと正則化項の勾配の大きさが各層で揃うため、層ごとに正則化の効き方が変わることなく、ネットワーク全体を効率的に学習することができる。最終層の正則化強度 λ_L は本発明の実施形態に係るニューラルネットワーク学習装置 100 への入力として指定される。これを調整することによって全ての層の学習を適切に行うことができる。正則化強度決定部 101 は、本例では比を一定とするよう

に正則化強度 λ_L を決定したが、差を一定にするように正則化強度 λ_L を決定しても良いし、一定でなく事前に定めた範囲内とするように正則化強度 λ_L を決定しても良い。また、基準となる正則化強度は、最終層でなくてもよく、どの層の強度でも良い。

[0030] ネットワーク学習部102は、学習データと初期化されたネットワークと正則化強度決定部101によって決定された正則化強度を用いてネットワークの学習を行う。ネットワーク学習部102は、学習には、一般に良く知られている誤差逆伝播法などを用いる。

[0031] 以上説明したように、本実施形態に係るニューラルネットワーク学習装置100においては、正則化強度決定部101が、初期化されたネットワークと基準の正則化強度とに基づいて層ごとの正則化強度を決定し、ネットワーク学習部102が、学習データと初期ネットワークと決定された正則化強度とに基づいてネットワークの学習を行い、学習したネットワークを出力する。

[0032] これにより、ニューラルネットワーク学習装置100は、ネットワークの構造に基づき、各層で適切な正則化強度を設定することができるので、ネットワーク全体を効率的に学習することができる。

[0033] 本実施形態の効果は、ネットワーク学習時に適切な正則化を行い、ネットワーク全体を効率よく学習できる点である。これは、層ごとに適切な正則化強度を設定しているためである。ネットワークの学習時、ネットワークのパラメータは、損失関数から計算されるパラメータ更新量と正則化項から計算されるパラメータ更新量に基づいて更新されるが、両者の大きさは層によって異なる。本実施形態の正則化強度決定部101は、損失関数による更新量の大きさと正則化項による更新量の大きさとの違いが各層で一定範囲内となるように正則化強度を決定する。これにより、図4の右列に示すように、損失関数による更新量の大きさと正則化項による更新量の大きさとのバランスが全ての層で揃うため、層によって正則化が強すぎたり弱すぎたりすることなく、ネットワーク全体を効率よく学習できる。

[0034] <本発明の実施形態の変形例>

次に、本発明の実施形態の変形例を説明する。

[0035] 上述の実施形態では、正則化強度決定部 101 は、損失関数の勾配の大きさと正則化項の勾配の大きさとの比を一定にするため、層ごとに正則化強度を決定したが、本変形例では、正則化強度決定部 101 は、逆に正則化強度を一定とし、損失関数の勾配に対して係数をかけても良い。この場合、ネットワーク学習部 102 は、正則化強度決定部 101 が算出した正則化強度の逆数を係数として用いる。

[0036] 本発明の実施形態によれば、画像処理や音声処理において、例えば顔認識や物体認識などのように、パターンを識別する用途に使用することができる。この場合は、ニューラルネットワーク学習装置を用いて学習したニューラルネットワークに基づいて認識を行うパターン認識装置を用いる。

[0037] また、上記の実施形態の一部又は全部は、以下の付記のようにも記載されるが、以下には限られない。

[0038] (付記 1)

初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定手段と、

初期化されたネットワークと前記正則化強度決定手段で決定された正則化強度とに基づいてネットワークの学習を行うネットワーク学習手段とを備え、

前記正則化強度決定手段は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定する

ことを特徴とするニューラルネットワーク学習装置。

[0039] (付記 2)

前記正則化強度決定手段は、損失関数の勾配の大きさと正則化項の勾配の大きさとの比が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記 1 に記載のニューラルネットワーク学習装置。

[0040] (付記3)

前記正則化強度決定手段は、損失関数の勾配の大きさと正則化項の勾配の大きさととの差が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記1に記載のニューラルネットワーク学習装置。

[0041] (付記4)

付記1乃至3のいずれか1項に記載のニューラルネットワーク学習装置を用いて学習したニューラルネットワークに基づいて認識を行うことを特徴とするパターン認識装置。

[0042] (付記5)

初期化されたネットワークに基づいて層ごとの正則化強度を決定し、初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行い、

損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさととの違いが所定範囲内となるように正則化強度を決定する

ことを特徴とするニューラルネットワーク学習方法。

[0043] (付記6)

損失関数の勾配の大きさと正則化項の勾配の大きさととの比が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記5に記載のニューラルネットワーク学習方法。

[0044] (付記7)

損失関数の勾配の大きさと正則化項の勾配の大きさととの差が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記5に記載のニューラルネットワーク学習方法。

[0045] (付記8)

付記5乃至7のいずれか1項に記載のニューラルネットワーク学習方法を用いて学習したニューラルネットワークに基づいて認識を行うことを特徴とするパターン認識方法。

[0046] (付記 9)

コンピュータに、

初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定処理と、

初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行う学習処理と、

を実行させ、

前記正則化強度決定処理は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定する、

ことを特徴とするプログラムを記憶する記憶媒体。

[0047] (付記 10)

前記正則化強度決定処理は、損失関数の勾配の大きさと正則化項の勾配の大きさとの比が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記 9 に記載の記憶媒体。

[0048] (付記 11)

前記正則化強度決定処理は、損失関数の勾配の大きさと正則化項の勾配の大きさとの差が所定範囲内となるように正則化強度を決定する

ことを特徴とする付記 9 に記載の記憶媒体。

[0049] 本発明は、上述の実施形態に限定されない。本発明は、本発明の趣旨を逸脱しない範囲において当業者が理解し得る様々な変更が可能である。

[0050] この出願は、2016年2月24日に提出された日本出願特願2016-032806を基礎とする優先権を主張し、その開示の全てをここに取り込む。

符号の説明

[0051] 100 ニューラルネットワーク学習装置
101 正則化強度決定部
102 ネットワーク学習部

200	コンピュータ装置
201	CPU
202	ROM
203	RAM
204	プログラム
205	記憶装置
206	記憶媒体
207	ドライブ装置
208	通信インターフェース
209	ネットワーク
210	入出力インターフェース
211	バス

請求の範囲

- [請求項1] 初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定手段と、
- 初期化されたネットワークと前記正則化強度決定手段で決定された正則化強度とに基づいてネットワークの学習を行うネットワーク学習手段とを備え、
- 前記正則化強度決定手段は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定することを特徴とするニューラルネットワーク学習装置。
- [請求項2] 前記正則化強度決定手段は、損失関数の勾配の大きさと正則化項の勾配の大きさとの比が所定範囲内となるように正則化強度を決定することを特徴とする請求項1に記載のニューラルネットワーク学習装置。
- [請求項3] 前記正則化強度決定手段は、損失関数の勾配の大きさと正則化項の勾配の大きさとの差が所定範囲内となるように正則化強度を決定することを特徴とする請求項1に記載のニューラルネットワーク学習装置。
- [請求項4] 請求項1乃至3のいずれか1項に記載のニューラルネットワーク学習装置を用いて学習したニューラルネットワークに基づいて認識を行うことを特徴とするパターン認識装置。
- [請求項5] 初期化されたネットワークに基づいて層ごとの正則化強度を決定し、
- 初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行い、
- 損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定する

ことを特徴とするニューラルネットワーク学習方法。

[請求項6] 損失関数の勾配の大きさと正則化項の勾配の大きさととの比が所定範囲内となるように正則化強度を決定する

ことを特徴とする請求項5に記載のニューラルネットワーク学習方法。

[請求項7] 損失関数の勾配の大きさと正則化項の勾配の大きさととの差が所定範囲内となるように正則化強度を決定する

ことを特徴とする請求項5に記載のニューラルネットワーク学習方法。

[請求項8] 請求項5乃至7のいずれか1項に記載のニューラルネットワーク学習方法を用いて学習したニューラルネットワークに基づいて認識を行うことを特徴とするパターン認識方法。

[請求項9] コンピュータに、
初期化されたネットワークに基づいて層ごとの正則化強度を決定する正則化強度決定処理と、

初期化されたネットワークと決定された前記正則化強度とに基づいてネットワークの学習を行う学習処理と、

を実行させ、

前記正則化強度決定処理は、損失関数から計算されるパラメータ更新量の大きさと正則化項から計算されるパラメータ更新量の大きさとの違いが所定範囲内となるように正則化強度を決定する、

ことを特徴とするプログラムを記憶する記憶媒体。

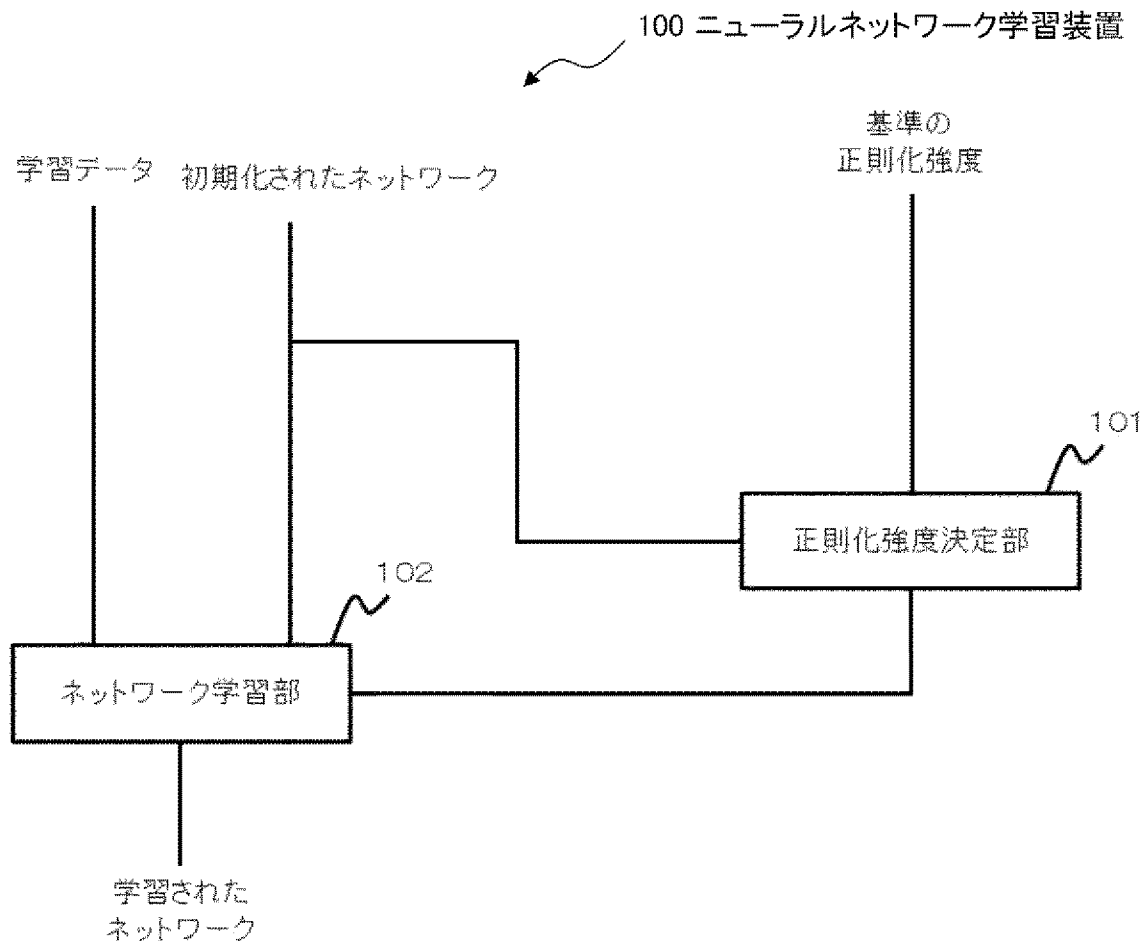
[請求項10] 前記正則化強度決定処理は、損失関数の勾配の大きさと正則化項の勾配の大きさととの比が所定範囲内となるように正則化強度を決定する

ことを特徴とする請求項9に記載の記憶媒体。

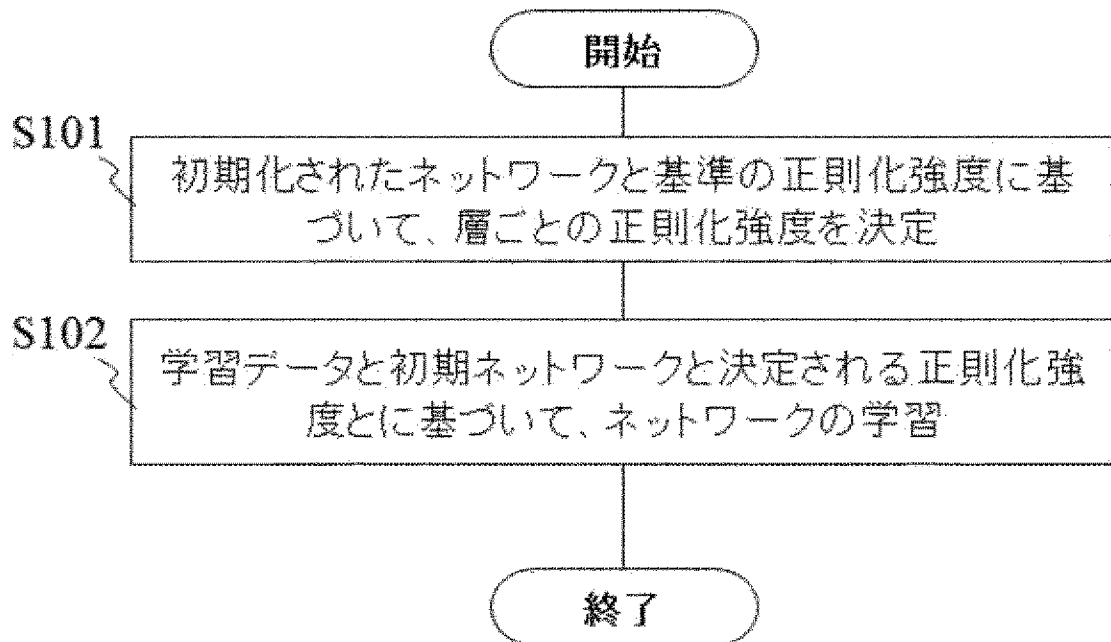
[請求項11] 前記正則化強度決定処理は、損失関数の勾配の大きさと正則化項の勾配の大きさととの差が所定範囲内となるように正則化強度を決定する

ことを特徴とする請求項9に記載の記憶媒体。

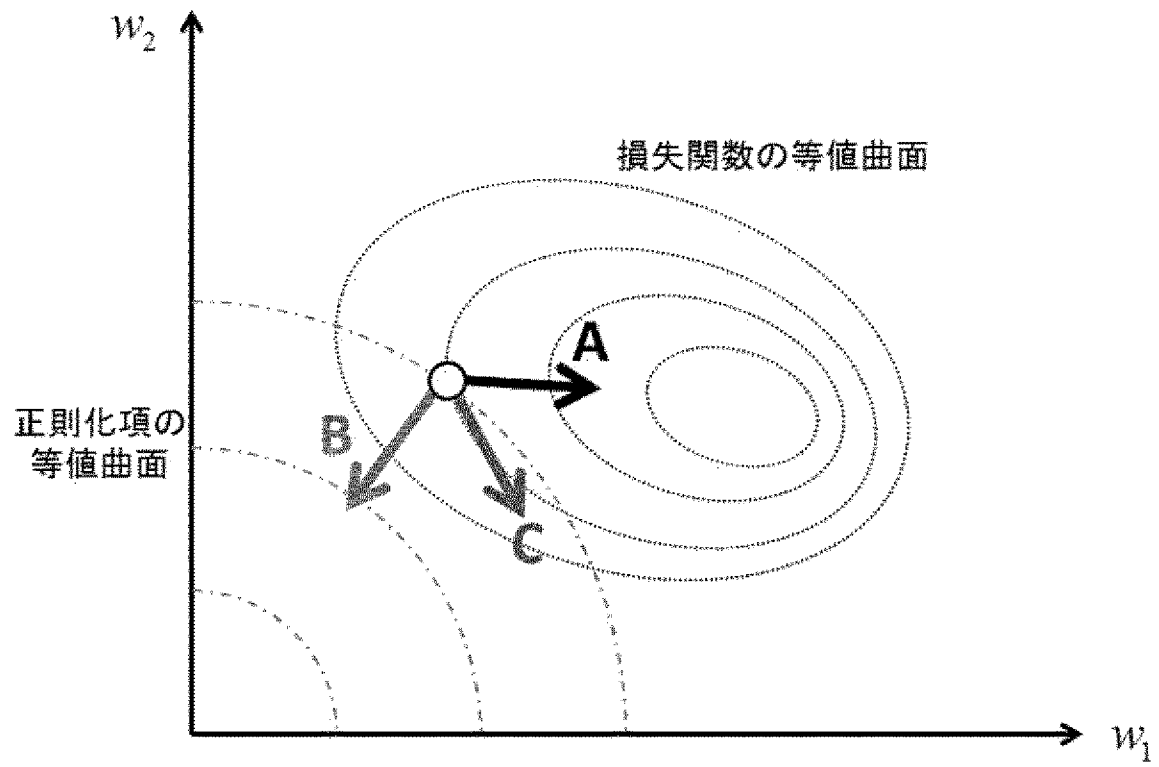
[図1]



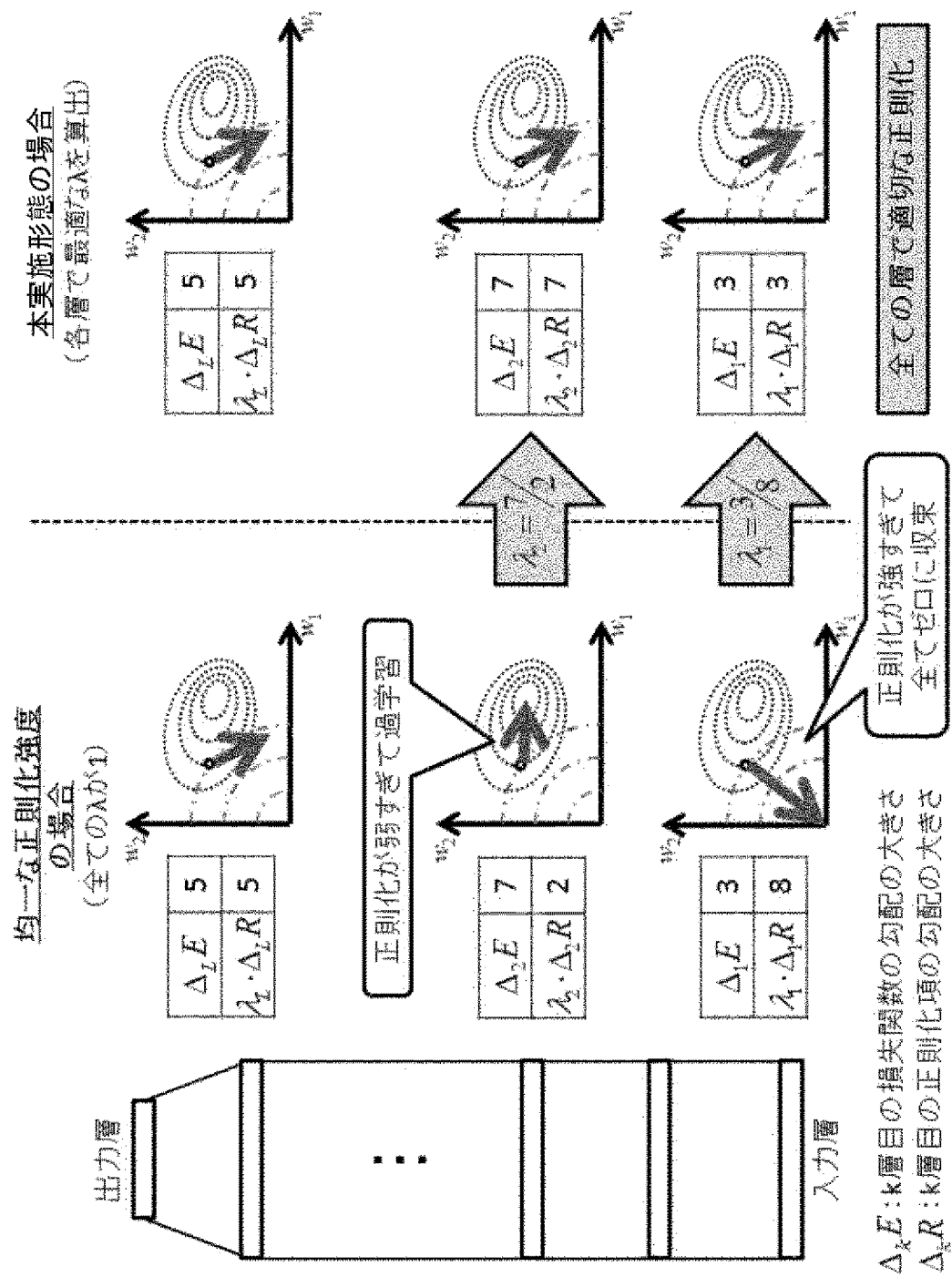
[図2]



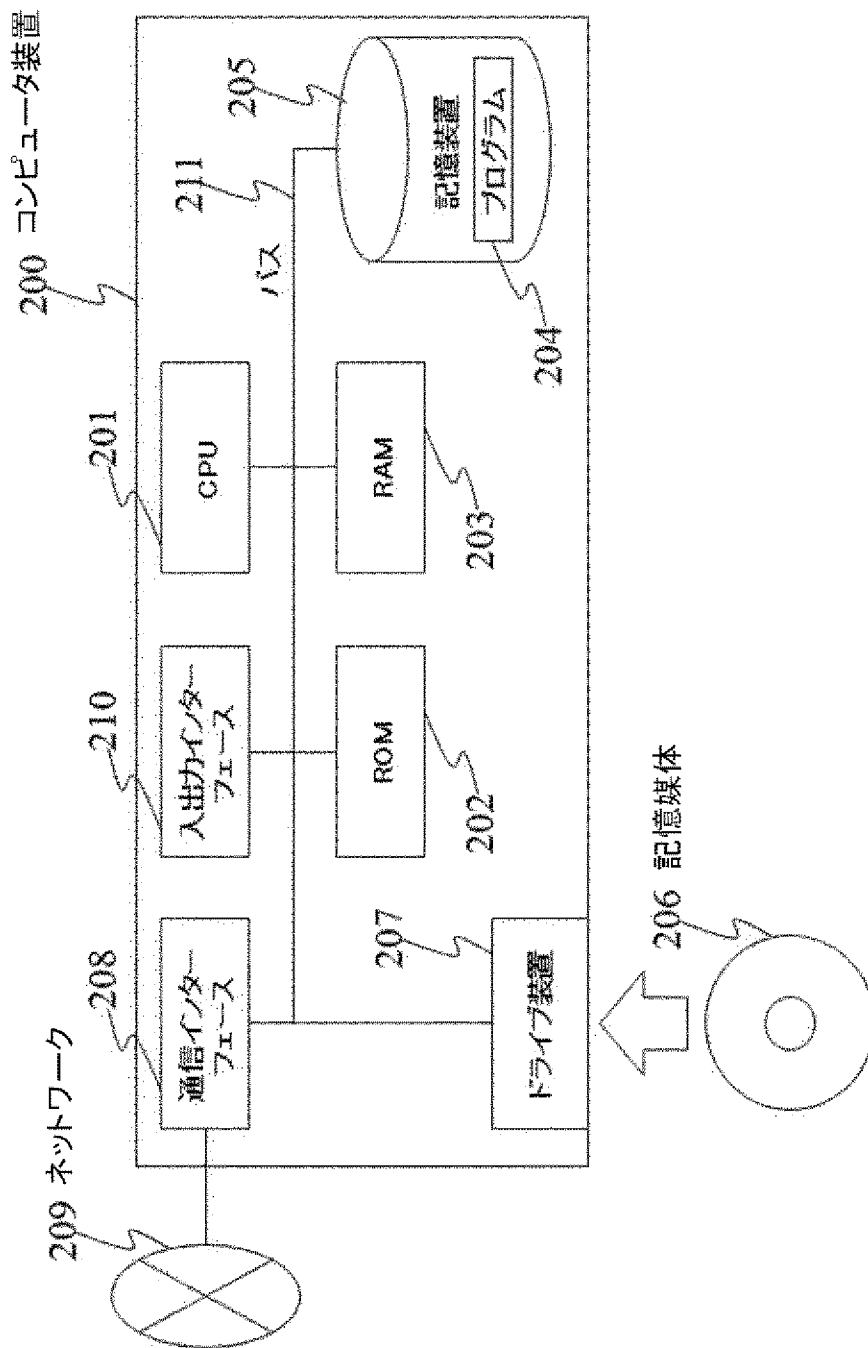
[図3]



[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/005230

A. CLASSIFICATION OF SUBJECT MATTER

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)

G06N3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

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	JP 8-202674 A (Ricoh Co., Ltd.), 09 August 1996 (09.08.1996), entire text; all drawings (Family: none)	1-11
A	JP 2001-142864 A (Nippon Telegraph and Telephone Corp.), 25 May 2001 (25.05.2001), entire text; all drawings (Family: none)	1-11
A	Kazumi SAITO, "A New Regularization Based on the MDL Principle", Journal of Japanese Society for Artificial Intelligence, 01 January 1998 (01.01.1998), vol.13, no.1, pages 123 to 130	1-11

☒ 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
05 April 2017 (05.04.17)

Date of mailing of the international search report
18 April 2017 (18.04.17)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/005230

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHI, Dung Doan, GENERALIZATION FOR MULTILAYER NEURAL NETWORK BAYESIAN REGULARIZATION OR EARLY STOPPING, Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference, Asia Pacific Association of Hydrology and Water Resources[online], 2004.07.08 [retrieval date 2017.03.28], Internet:<URL: http://rwes.dpri.kyoto-u.ac.jp/~tanaka/APHW/APHW2004/proceedings/FWR/56-FWR-M185/56-FWR-M185%20(1).pdf >	1-11

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. G06N3/08

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	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017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	JP 8-202674 A（株式会社リコー）1996.08.09, 全文全図（ファミリーなし）	1-11
A	JP 2001-142864 A（日本電信電話株式会社）2001.05.25, 全文全図（ファミリーなし）	1-11
A	斉藤 和巳, MDL原理に基づく新正則化法, 人工知能学会誌, 1998.01.01, 第13巻/第1号, pp.123-130	1-11

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

05.04.2017

国際調査報告の発送日

18.04.2017

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）

塚田 肇

5 B

3652

電話番号 03-3581-1101 内線 3545

C (続き) . 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	CHI, Dung Doan, GENERALIZATION FOR MULTILAYER NEURAL NETWORK BAYESIAN REGULARIZATION OR EARLY STOPPING, Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference, Asia Pacific Association of Hydrology and Water Resources [オンライン], 2004.07.08 [検索日 2017.03.28], インターネット:< URL:http://rwes.dpri.kyoto-u.ac.jp/~tanaka/APHW/APHW2004/proceedings/FWR/56-FWR-M185/56-FWR-M185%20(1).pdf>	1 — 1 1