

(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
2018年11月1日(01.11.2018)



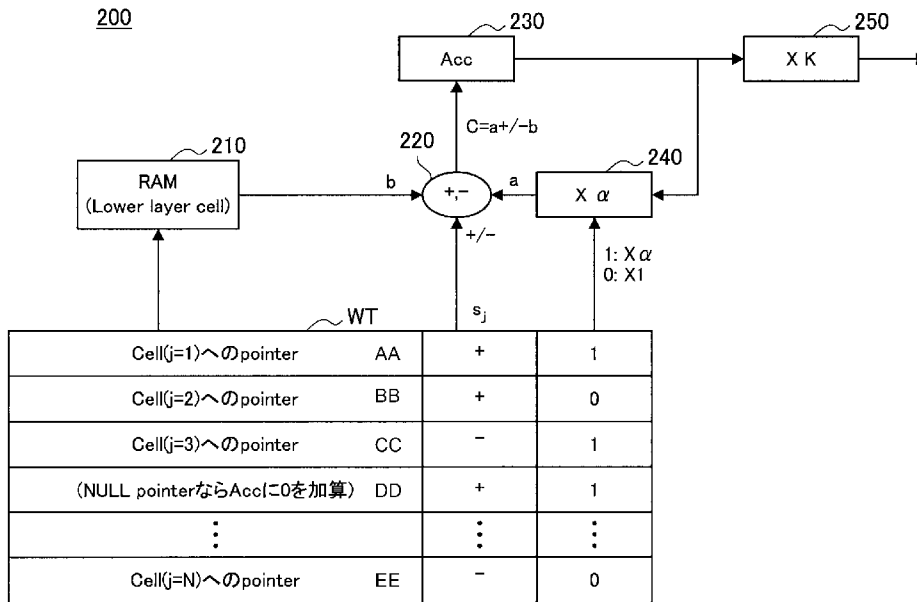
(10) 国際公開番号
WO 2018/198454 A1

- (51) 国際特許分類:
G06F 17/10 (2006.01) G06N 99/00 (2010.01)
G06N 3/063 (2006.01)
- (21) 国際出願番号: PCT/JP2018/002404
- (22) 国際出願日: 2018年1月26日(26.01.2018)
- (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 2017-089144 2017年4月28日(28.04.2017) JP
- (71) 出願人: ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 廣井 聡幸 (HIROI, Toshiyuki); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP). 山本 真紀子 (YAMAMOTO, Makiko); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP). 中村 章 (NAKAMURA, Akira); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (74) 代理人: 亀谷 美明, 外 (KAMEYA, Yoshiaki et al.); 〒1600004 東京都新宿区四谷3-1-3 第一冨澤ビル はづき国際特許事務所四谷オフィス 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,

(54) Title: INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称: 情報処理装置、および情報処理方法

[図8]



AA... Pointer TO cell (j=1)
 BB... Pointer TO cell (j=2)
 CC... Pointer TO cell (j=3)
 DD... ADD 0 TO Acc IF NULL pointer
 EE... Pointer TO cell (j=N)

(57) Abstract: [Problem] To further lighten a processing load associated with inner-product computation and ensure the quantization granularity of a weighting coefficient. [Solution] Provided is an information processing device equipped with a computation unit for performing inner-product computation based on a plurality of input values and a plurality of weighting coefficients that respectively correspond to the input values and calculating an output value, the computation unit calculating the output value on the basis of the weighting coefficients quantized on the basis of granularity in the

WO 2018/198454 A1

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, 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))

vector direction on a N-dimensional hyperspherical surface. The present invention is also equipped with a sum-of-product computation circuit for executing sum-of-product computation based on a plurality of input values and a plurality of weighting coefficients that respectively correspond to the input values, the sum-of-product computation circuit being provided with: a storage circuit for holding the address information of the input values, which are sorted in increasing order of values and correspond to the plurality of weighting coefficients; and a multiplication circuit for executing computation based on the weighting coefficients and the input values acquired by the address information.

- (57) 要約: 【課題】内積演算に係る処理負担をより軽減すると共に、重み係数の量子化粒度を保証する。
 【解決手段】複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく内積演算を行い出力値を算出する演算部、を備え、前記演算部は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出する、情報処理装置が提供される。また、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく積和演算を実行する積和演算回路を備え、前記積和演算回路は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持する記憶回路と、前記アドレス情報により取得される前記入力値と前記重み係数に基づく演算を実行する乗算回路と、を備える、情報処理装置が提供される。

明 細 書

発明の名称：情報処理装置、および情報処理方法

技術分野

[0001] 本開示は、情報処理装置、および情報処理方法に関する。

背景技術

[0002] 近年、脳神経系の仕組みを模した数学モデルであるニューラルネットワークが注目されている。また、ニューラルネットワークにおける演算の処理負担を軽減するための種々の手法が提案されている。例えば、非特許文献1には、重み係数を2値化することで、処理負担を軽減する手法が記載されている。また、非特許文献2には、入力信号をログドメインに変換することにより、乗算を加算に変換する手法が記載されている。

先行技術文献

非特許文献

[0003] 非特許文献1：Matthieu Courbariaux、外2名、“BinaryConnect: Training Deep Neural Networks with binary weights during propagations”、[online]、2015年11月2日、arXiv、[平成29年3月22日検索]、インターネット〈URL: <https://arxiv.org/pdf/1511.00363.pdf>〉

非特許文献2：Daisuke Miyashita、外2名、“Convolutional Neural Networks using Logarithmic Data Representation”、[online]、2016年3月3日、arXiv、[平成29年3月22日検索]、インターネット〈URL: <https://arxiv.org/pdf/1603.01025.pdf>〉

発明の概要

発明が解決しようとする課題

[0004] しかし、非特許文献1に記載の手法では、+1または-1を用いた2値化を行っているため、重み係数の次元数が増加するにつれて量子化の粒度が荒くなることが想定される。また、非特許文献2に記載の手法では、乗算の回避において所定の効果を有するものの、処理負担の軽減については、さらな

る余地があるものと想定される。

[0005] そこで、本開示では、内積演算に係る処理負担をより軽減すると共に、重み係数の量子化粒度を保証することが可能な、新規かつ改良された情報処理装置および情報処理方法を提案する。

課題を解決するための手段

[0006] 本開示によれば、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく内積演算を行い出力値を算出する演算部、を備え、前記演算部は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出する、情報処理装置が提供される。

[0007] また、本開示によれば、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく積和演算を実行する積和演算回路を備え、前記積和演算回路は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持する記憶回路と、前記アドレス情報により取得される前記入力値と前記重み係数に基づく演算を実行する乗算回路と、を備える、情報処理装置が提供される。

[0008] また、本開示によれば、プロセッサが、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく積和演算を行い出力値を算出すること、を含み、前記算出することは、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出すること、をさらに含む、情報処理方法が提供される。

発明の効果

[0009] 以上説明したように本開示によれば、内積演算に係る処理負担をより軽減すると共に、重み係数の量子化粒度を保証することが可能となる。

[0010] なお、上記の効果は必ずしも限定的なものではなく、上記の効果とともに、または上記の効果に代えて、本明細書に示されたいずれかの効果、または本明細書から把握され得る他の効果が奏されてもよい。

図面の簡単な説明

[0011] [図1]本開示の関連技術に係るニューラルネットワークにおける基本演算の概要について説明するための概念図である。

[図2]本開示の関連技術に係る入力ベクトルと重みベクトルとの内積演算について説明するための概要図である。

[図3]本開示の関連技術に係る2次元空間において2値量子化した重みベクトルについて説明するための図である。

[図4]本開示の関連技術に係る2次元空間において4値量子化した重みベクトルについて説明するための図である。

[図5]本開示の関連技術に係る3次元空間における重みベクトルの粒度のばらつきについて説明するための図である。

[図6]本開示の関連技術に係るN次元空間における重みベクトルの粒度のばらつきについて説明するための図である。

[図7]本開示の一実施形態に係る情報処理装置の機能ブロック図の一例である。

[図8]同実施形態に係る情報処理装置が備える積和演算回路の回路ブロック図の一例である。

[図9]同実施形態に係るアドレステーブルが保持するアドレス情報に係るオフセット表記の一例である。

[図10]同実施形態に係る情報処理方法の処理イメージを示す図である。

[図11]同実施形態に係る量子化粒度 $\Delta\theta$ について説明するための図である。

[図12]同実施形態に係る α に応じた量子化粒度 $\Delta\theta$ の最大値を示すグラフである。

[図13]同実施形態に係る最大べき乗数について説明するための図である。

[図14]同実施形態に係る入力数Nに対する乗算回数の例を示す図である。

[図15]同実施形態に係る入力数Nに対する乗算回数の例を示す図である。

[図16]本開示の一実施形態に係るハードウェア構成例を示す図である。

発明を実施するための形態

[0012] 以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細

に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

[0013] なお、説明は以下の順序で行うものとする。

1. 実施形態

1. 1. 背景

1. 2. 情報処理装置10の機能構成例

1. 3. 重みベクトルの量子化

1. 4. 積和演算回路の構成例

1. 5. 学習時における量子化

1. 6. 効果

2. ハードウェア構成例

3. まとめ

[0014] <1. 実施形態>

<<1. 1. 背景>>

近年、深層学習 (Deep Learning) などニューラルネットワークを用いた学習手法が広く研究されている。ニューラルネットワークを用いた学習手法は高い精度を有する一方、演算に係る処理負担が大きいことから、当該処理負担を効果的に軽減する演算方式が求められている。

[0015] このため、近年においては、例えば、Binary weight networkや、XNOR netなどの、演算量や情報量の軽減を目的とする演算方式も提案されている。

[0016] ここで、一般的なニューラルネットワークにおける基本演算の概要について説明する。図1は、ニューラルネットワークにおける基本演算の概要について説明するための概念図である。図1には、ニューラルネットワークを構成する2つの層と、当該2つの層にそれぞれ属するセル $c_{1_1} \sim c_{1_N}$ とセル c_{2_1} とが示されている。

[0017] この際、セル C_{2_1} に入力される入力信号（以下、入力ベクトル、とも称す

る)は、下層に属するセル $c_{1_1} \sim c_{1_N}$ に係る入力ベクトルと重み係数(以下、重みベクトル、とも称する)とに基づいて決定される。より詳細には、セル c_{2_1} に入力される入力ベクトルは、セル $c_{1_1} \sim c_{1_N}$ に係る入力ベクトルと重みベクトルとの内積演算結果にバイアス b を加算し、さらに活性化関数 h により処理した値となる。

[0018] すなわち、セル $c_{1_1} \sim c_{1_N}$ に係る入力ベクトルと重みベクトルとをそれぞれ、 $x = (x_1, x_2, x_3, \dots, x_N)$ 、 $w = (w_1, w_2, w_3, \dots, w_N)$ とした場合、セル c_{2_1} に入力される入力ベクトル z は、下記の数式(1)により定義される。

[0019] [数1]

$$z = h\left(\sum x_i w_i + b\right) \quad \dots (1)$$

[0020] この際、重みベクトル w を法線とする超平面 h を想定すると、 N 次元で表現される入力ベクトル x と重みベクトル w との内積は、図2に示すように、超平面 h を基準とした入力ベクトル x を重みベクトル w へ射影した射影距離 d に $\|w\|$ を掛けた値により表現することができる。ここで、上記の $\|w\|$ は、ベクトル w のノルムである。図2は、入力ベクトル x と重みベクトル w との内積演算について説明するための概要図である。

[0021] ここで、入力ベクトル x と重みベクトル w とが2次元である場合、非特許文献1のように、重みベクトル w を $+1$ または -1 の2値で量子化すると、重みベクトル w は、図3のように表すことができる。図3は、2次元空間において2値量子化した重みベクトル w について説明するための図である。この際、重みベクトル w の粒度は、平面における回転角度 θ で表現することができ、当該粒度は、図3に示すように、 90 度となる。

[0022] 続いて、重みベクトル w を $\{0, 1/4, 1/2, 1\}$ の4値で量子化する場合を考察する。図4は、2次元空間において4値量子化した重みベクトル w について説明するための図である。この場合、重みベクトル w の粒度、すなわち回転角度 θ は、約 15 度となり、2値量子化した場合と比較してよ

り細かい粒度を保証することが可能となる。

[0023] 一方、重みベクトル w の粒度は、次元数が増加するにつれ、ばらつきが大きくなることが想定される。図5は、3次元空間における重みベクトル w の粒度のばらつきについて説明するための図である。例えば、図5に示すように、 $(1, 1, 0)$ と $(0, 0, 1)$ とで定義される面における重みベクトル w の量子化を考えた場合、 $(1, 1, 0)$ 方向の辺の長さは、 $(0, 0, 1)$ 方向の辺の長さの $\sqrt{2}$ 倍となるため、量子化時における粒度のばらつきが大きくなることがわかる。

[0024] また、上記のような粒度のばらつき、すなわち不均一性は、次元数が増加するほど顕著となる。図6は、 N 次元空間における重みベクトル w の粒度のばらつきについて説明するための図である。図6には、 N 次元空間において、 $(1, 1, \dots, 1, 0)$ と $(0, 0, \dots, 0, 1)$ で定義される面が示されている。この際、 $(1, 1, \dots, 1, 0)$ 方向の辺の長さは、 $(0, 0, \dots, 0, 1)$ 方向の辺の長さの $\sqrt{N-1}$ 倍により表すことができる。例えば、 $N=100$ である場合には、 $(1, 1, \dots, 1, 0)$ 方向の辺の長さは、 $(0, 0, \dots, 0, 1)$ 方向の辺の $\sqrt{99}$ 倍（ ≈ 10 倍）となる。

[0025] 上記のような重みベクトル w に係る粒度の不均一性は、ニューラルネットワークの性能低下の要因となり得ることから、より精度の高い量子化手法が望まれていた。本開示に係る技術思想は、上記の点に着目して発想されたものであり、 N 次元空間における重みベクトルの高い近似精度を維持しながらも効果的に処理負担を軽減することを可能とする。このために、本開示の一実施形態に係る情報処理装置および情報処理方法は、 N 次元超球平面におけるベクトル方向の粒度に基づいて量子化された重みベクトルを用いて内積演算を行うことを特徴の一つとする。本開示の一実施形態に係る情報処理装置および情報処理方法は、重みベクトルを、細かすぎず、かつ粗すぎない粒度で量子化することで、高い近似精度と処理負担の軽減を両立することが可能である。より詳細には、本開示の一実施形態に係る情報処理装置および情報

処理方法は、べき乗により表現された重みベクトルを用いた内積演算を行ってよい。以下、本開示の一実施形態に係る情報処理装置および情報処理方法の有する上記の特徴について詳細に説明する。

[0026] <<1. 2. 情報処理装置10の機能構成例>>

続いて、本実施形態に係る情報処理方法を実現する情報処理装置10の機能構成例について述べる。図7は、本実施形態に係る情報処理装置10の機能ブロック図の一例である。図7を参照すると、本実施形態に係る情報処理装置10は、入力部110、演算部120、記憶部130、および出力部140を備える。以下、上記構成について、当該構成が有する機能を中心に説明する。

[0027] (入力部110)

本実施形態に係る入力部110は、操作者による種々の入力操作を検出する機能を有する。このために、本実施形態に係る入力部110は、操作者による入力操作を検出するための各種の装置を含んでよい。入力部110は、例えば、各種のボタン、キーボード、タッチパネル、マウス、スイッチなどにより実現され得る。

[0028] (演算部120)

本実施形態に係る演算部120は、複数の入力値と当該入力値にそれぞれ対応する複数の重み係数とに基づく内積演算を行い出力値を算出する機能を有する。本実施形態に係る演算部120は、特に、ニューラルネットワークの順伝播に係る内積演算を行う。この際、本実施形態に係る演算部120は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された重み係数に基づいて出力値を算出することを特徴の一つとする。より具体的には、本実施形態に係る演算部120は、べき乗により表現された重み係数に基づいて出力値を算出してよい。本実施形態に係る内積演算の特徴については、別途詳細に説明する。

[0029] (記憶部130)

記憶部130は、情報処理装置10が備える各構成で用いられるプログラ

ムやデータなどを記憶する機能を有する。本実施形態に係る記憶部130は、例えば、ニューラルネットワークに用いられる種々のパラメータなどを記憶する。

[0030] (出力部140)

出力部140は、操作者に対し種々の情報出力を行う機能を有する。このために、本実施形態に係る出力部140は、視覚情報を出力するディスプレイ装置を含んで構成され得る。ここで、上記のディスプレイ装置は、例えば、CRT (Cathode Ray Tube) ディスプレイ装置、液晶ディスプレイ (LCD: Liquid Crystal Display) 装置、OLED (Organic Light Emitting Diode) 装置などにより実現され得る。

[0031] 以上、本実施形態に係る情報処理装置10の機能構成例について説明した。なお、上記で説明した機能構成例はあくまで一例であり、本実施形態に係る情報処理装置10の機能構成例は係る例に限定されない。本実施形態に係る情報処理装置10は、図1に示す以外の構成をさらに備えてもよい。情報処理装置10は、例えば、他の情報処理端末との情報通信を行う通信部などをさらに備えてもよい。本実施形態に係る情報処理装置10の機能構成は、柔軟に設計変更され得る。

[0032] <<1. 3. 重みベクトルの量子化>>

次に、本実施形態に係る重みベクトルの量子化について詳細に説明する。上述したように、本実施形態に係る情報処理装置10は、べき乗により表現された重みベクトル w による量子化を行うことで、粒度の均一性を高く保つことができる。この際、本実施形態に係る演算部120は、複数の重みベクトル w_i を値の小さい順番に並び替え、また当該複数の重みベクトル w_i を最も値の大きい重み係数 w_i で正規化することを特徴の一つとする。ここで、並び替えて正規化された重みベクトルを w_j とすると、重みベクトル w_j は、下記の数式(2)～(4)により表される。

[0033]

[数2]

$$w_j = s_j \alpha^{n_j} \quad (j < N) \quad \dots (2)$$

$$w_N = 1 \quad \dots (3)$$

$$w_j \leq w_{j+1} \quad \dots (4)$$

[0034] 但し、この際、上記の数式(2)における α は、 $0 < \alpha < 1$ 、であり、 s_j は、 $s_j \in \{-1, 1\}$ であり、 n_j は、 $n_j \in \{0, 1, 2, \dots\}$ 、であってよい。すなわち、本実施形態に係る演算部120は、 n_j を整数として量子化を行う。この際、演算部120が実行する内積演算は、下記の数式(5)により表される。なお、下記の数式(5)における K は、正規化定数を示す。また、上記 α の値については、下記の数式(5)を適宜変形した場合であっても、内積演算において最終的に上記の範囲に定まればよい。本開示に示す数式は、あくまで一例であり、柔軟に変形され得る。

[0035] [数3]

$$K((\dots(s_1 x_1 \alpha^{n_1 - n_2} + s_2 x_2) \alpha^{n_2 - n_3} + s_3 x_3) \alpha^{n_3 - n_4} + \dots) \alpha^{n_{N-1}} + s_N x_N),$$

$$K \in R \quad \dots (5)$$

[0036] このため、本実施形態に係る演算部120による内積演算は、 N 回の加算演算と、 $-1 / 2 \log(N-1) / \log \alpha$ のオーダーの乗算回数により処理が可能であることが予想される。

[0037] このように、本実施形態に係る情報処理方法では、重みベクトル w を α のべき乗表現により近似し、値の小さい順番に重みベクトル w を並び替えることを特徴の一つとする。この際、本実施形態に係る情報処理方法では、 N に応じて α の指数を k 値化することで、重みベクトル w の量子化が行われる。

[0038] 例えば、 $N = 100$ である場合、本実施形態に係る情報処理方法では、 $k = 4$ (2 bit), 8 (3 bit), 16 (4 bit) などが採用されてもよい。上記のように k を設定することで、上記の数式(5)における $n_1 - n_2$, $n_2 - n_3$, $n_3 - n_4, \dots$ の多くは、同一の値により量子化されることで0となるため、乗算の回数を大幅に削減することが可能となる。より具体的な

例を挙げると、 $N = 100$ に対し、 $k = 4$ である場合、 $n_{j-1} - n_j$ が0以外の値を取り得るのは4回のみとなる。このため、本例の場合、内積演算に係る乗算回数は4回で済み、残りはすべて加算となるため、処理負担を効果的に軽減することができる。

[0039] <<1. 4. 積和演算回路の構成例>>

次に、本実施形態に係る演算方式を実現する積和演算回路について説明する。上述のように、重みベクトル w をべき乗表現により量子化し、また並び替えを行った場合、重みベクトル w に対応する入力ベクトル x の並び替えも併せて行う必要がある。

[0040] このため、本実施形態に係る情報処理装置10は、値の小さい順番に並び替えられた複数の重みベクトル w と対応する入力ベクトル x のアドレス情報を保持するテーブルを有する積和演算回路を備えてよい。

[0041] 図8は、本実施形態に係る情報処理装置10が備える積和演算回路200の回路ブロック図の一例である。図8に示すように、本実施形態に係る積和演算回路は、重みベクトル w と対応する入力ベクトル x のアドレス情報を保持するテーブル WT を保持する記憶回路と、RAM210、加算回路220、アキュムレータ230、 α に係る乗算を行う第1の乗算回路240、および正規化定数に係る乗算を行う第2の乗算回路250を備える。

[0042] (アドレステーブル WT)

本実施形態に係るアドレステーブル WT は、値の小さい順番に並び替えられた複数の重みベクトル w と対応する入力ベクトル x のアドレス情報、符号情報、および乗算指示情報を保持する。なお、上記のアドレス情報は、図8に示すように、Null Pointerを含んでもよい。この場合、アキュムレータ230には0が加算されることとなり、アキュムレータ230の値を単純に α 倍することが可能となる。また、上記の符号情報は、上述した数式(5)における S_j に対応する値を示す情報である。

[0043] また、上記の乗算指示情報は、第1の乗算回路240による処理内容を指示する情報である。本実施形態に係る乗算指示情報は、例えば、乗算の可否

を指定する情報を含んでよい。図8には、乗算指示情報が0である場合には、第1の乗算回路240は乗算を行わず、乗算指示情報が1である場合には、第1の乗算回路240は α を乗算する場合の一例が示されている。

[0044] なお、本実施形態に係る乗算指示情報は、上記の例に限定されず、種々の処理内容を指定する情報を含んでよい。本実施形態に係る乗算指示情報は、例えば、乗算の回数やシフト演算を指定する情報などを含み得る。

[0045] (RAM210)

本実施形態に係るRAM210は、アドレステーブルWTから入力されたアドレス情報に基づいて重みベクトル w に対応する入力ベクトル x を加算回路220に出力する。

[0046] (加算回路220)

本実施形態に係る加算回路220は、RAM210から入力される入力ベクトル x と第1の乗算回路240から出力される値に基づいて加算を実行する。この際、本実施形態に係る加算回路220は、アドレステーブルWTが保持する符号情報に基づいて上記の加算を行う。

[0047] (アキュムレータ230)

本実施形態に係るアキュムレータ230は、加算回路220から出力される演算結果を累積する。アキュムレータ230は、累積した値を第1の乗算回路240および第2の乗算回路250に出力する。また、アキュムレータ230には、累積した値を0にリセットするためのリセット信号が適宜入力される。

[0048] (第1の乗算回路240)

本実施形態に係る第1の乗算回路240は、アキュムレータ230が累積する値に対し、 α を乗算する。この際、第1の乗算回路240は、上述したように、アドレステーブルWTが保持する乗算指示情報に基づいて上記の乗算を実行する。第1の乗算回路240は、演算結果を加算回路220に出力する。

[0049] (第2の乗算回路250)

本実施形態に係る第2の乗算回路250は、アキュムレータ230から出力される値に対し、正規化定数Kを乗算する。

[0050] 以上、本実施形態に係る積和演算回路200の構成例について説明した。本実施形態に係る積和演算回路200によれば、内積演算における乗算回数を効果的に削減し、処理負担を軽減することが可能となる。

[0051] 一方、本実施形態に係るアドレステーブルWTは、入力ベクトルxのアドレス情報を保持することで容量が大きくなるため、消費電力が増加することも想定される。このため、本実施形態に係るアドレステーブルWTは、図9に示すように、アドレス間の相対位置を示すオフセットを含んでもよい。図9は、本実施形態に係るアドレステーブルWTが保持するアドレス情報に係るオフセット表記の一例である。

[0052] 図9に示すように、本実施形態に係るアドレステーブルWTは、上述した数式(5)において $n_{j-1} - n_j$ の値が連続して0である区間、すなわち乗算が行われない区間のアドレスをアドレス順にソートし、当該アドレス間のオフセットをアドレス情報として保持してもよい。本実施形態に係る上記のアドレステーブルWTによれば、アドレス情報に係る情報量を大幅に削減し、消費電力を効果的に低減することが可能となる。

[0053] なお、本実施形態に係るアドレステーブルWTは、図8および図9に示した形式以外にも種々の形態を取り得る。本実施形態に係るアドレステーブルWTは、例えば、符号情報と乗算指示情報とを明確に分離して保持しなくてもよいし、上記以外のアドレス圧縮方法を採用してもよい。本実施形態に係るアドレステーブルWTは、ニューラルネットワークの構成や情報処理装置10の性能などに応じて柔軟に変形され得る。

[0054] <<1.5. 学習時における量子化>>

次に、本実施形態に係る学習時における重みベクトル w_i の量子化について説明する。本実施形態に係る情報処理方法において、学習時における重みベクトル w_i の更新については、下記の数式(6)により計算することが可能である。

[0055] [数4]

$$n_i = \text{int}(\log|w_i/w_{\max}|/\log\alpha) \quad \dots (6)$$

[0056] 但し、上記の数式(6)における w_{\max} は、 w_i の最大値を示す。また、整数化 int については、切り上げ、または切り捨てのうちいずれか近い方を選んでよい。本実施形態に係る情報処理方法では、最終学習時に、 n_i を並び替えることで、上述したアドレステーブルWTを生成することができる。

[0057] なお、1に近い α を設定した場合においては、通常の浮動小数点演算を用いたDNN(Deep Neural Network)による学習アルゴリズムを用いて演算を行い、学習終了後の w_i を量子化する手法を採択することも可能である。すなわち、重みベクトル w に係る量子化粒度が小さい場合には、ネットワーク構造を変形しなくても、認識精度が低下しない効果が期待される。

[0058] <<1.6. 効果>>

次に、本実施形態に係る重みベクトルの量子化により奏される効果について詳細に説明する。上述したように、本実施形態に係る情報処理方法では、べき乗表現により量子化した w_i を値の小さい順番に並び替え正規化を行うことで w_j と定義する。この際、並び替えられた基底ベクトルを q_j とすると、重みベクトル w は、下記の数式(7)により表される。

[0059] [数5]

$$\begin{aligned} w &= \sum w_j q_j \\ &= ((\dots(s_1 q_1 \alpha^{n_1-n_2} + s_2 q_2) \alpha^{n_2-n_3} + s_3 q_3) \alpha^{n_3-n_4} + \dots) \alpha^{n_{N-1}} + s_N q_N) \quad \dots (7) \end{aligned}$$

[0060] すなわち、本実施形態に係る情報処理方法は、図10に示すように、 q_1, q_2, \dots, q_{j-1} で張られる空間に投影された重みベクトルと q_j とにより形成される平面にベクトルを作成し、当該ベクトルを $\alpha^{n_j-n_{j+1}}$ 倍する処理を繰り返す、と意味を有する。図10は、本実施形態に係る情報処理方法の処理イメージを示す図である。

[0061] このため、本実施形態に係る情報処理方法では、重みベクトルを q_1, q_2, \dots, q_{j-1} 空間に投影した軸と q_j とで張られる平面において、重みベクトルの量子化粒度 $\Delta\theta$ は、図 11 に示すように、反時計回転および時計回転方向にそれぞれ下記の数式 (8) および (9) により表すことができる。但し、この際、数式 (8) および (9) における l は、数式 (10) により定義される。図 11 は、本実施形態に係る量子化粒度 $\Delta\theta$ について説明するための図である。なお、図 11 では、第 1 象限に射影された重みベクトルが示されている。

[0062] [数6]

$$\Delta\theta_1 = \tan^{-1} \frac{1}{\alpha l} - \tan^{-1} \frac{1}{l} \quad \dots (8)$$

$$\Delta\theta_2 = \tan^{-1} \frac{1}{l} - \tan^{-1} \frac{\alpha}{l} \quad \dots (9)$$

$$l = \left\| \left(\dots \left((s_1 q_1 \alpha^{n_1 - n_2} + s_2 q_2) \alpha^{n_2 - n_3} + s_3 q_3 \right) \alpha^{n_3 - n_4} + \dots q_{j-1} \right) \alpha^{n_{j-1} - n_j} \right\| \quad \dots (10)$$

[0063] また、量子化粒度 $\Delta\theta_1$ および $\Delta\theta_2$ を l で微分して 0 とおいた場合、それぞれの最大値は、下記の数式 (11) により定義される。

[0064] [数7]

$$\Delta\theta_{1\max} = \Delta\theta_{2\max} = \tan^{-1} \frac{1}{\sqrt{\alpha}} - \tan^{-1} \sqrt{\alpha} \quad \dots (11)$$

[0065] 図 12 は、本実施形態に係る α に応じた量子化粒度 $\Delta\theta$ の最大値を示すグラフである。このように、本実施形態に係る情報処理方法によれば、N次元空間において、すべての直交する回転方向に量子化粒度が保証される。

[0066] なお、べき乗演算を途中で停止した場合、図 13 に示すように、量子化粒度 $\Delta\theta$ が残されることとなる。図 13 は、本実施形態に係る最大べき乗数について説明するための図である。なお、図 13 では、第 1 象限に射影された重みベクトルが示されている。この際、量子化粒度 $\Delta\theta$ を保証する最大べき乗数は、下記の数式 (12) を満たす最小な m に対し、下記の数式 (13)

を加算すればよい。従って、本実施形態に係る情報処理装置 10 が実行する乗算回数は、下記の数式 (14) により求めることができる。

[0067] [数8]

$$\tan^{-1} \alpha^m < \tan^{-1} \frac{1}{\sqrt{\alpha}} - \tan^{-1} \sqrt{\alpha} \quad \dots (12)$$

$$\frac{1 \log(N-1)}{2 \log \alpha} \quad \dots (13)$$

$$\frac{\log(\tan(\tan^{-1} \frac{1}{\sqrt{\alpha}} - \tan^{-1} \sqrt{\alpha}))}{\log \alpha} - \frac{1 \log(N-1)}{2 \log \alpha} + 1 \quad \dots (14)$$

[0068] 例えば、 $\alpha = 3/4$ である場合、入力数 N に対する乗算回数は、図 14 に示すグラフのように決定される。また、例えば、 $\alpha = 7/8$ である場合には、入力数 N に対する乗算回数は、図 15 に示すグラフのように決定される。すなわち、本実施形態に係る演算部 120 による内積演算において、乗算回数は、重みベクトルの底の値に基づいて定まるといえる。図 14 および図 15 は、本実施形態に係る入力数 N に対する乗算回数の例を示す図である。

[0069] 以上説明したように、本実施形態に情報処理方法を実現する情報処理装置によれば、ニューラルネットワークの順伝播に係る内積演算において、乗算回数を大幅に削減することができ、積和演算回路 200 による消費電力を効果的に低減することが可能となる。また、本実施形態に情報処理方法を実現する情報処理装置によれば、重みベクトルの量子化精度を向上させることができ、同一のビット数による従来の量子化手法と比較して、ニューラルネットワークによる認識精度および近似精度を向上させる効果が期待される。

[0070] <2. ハードウェア構成例>

次に、本開示の一実施形態に係る情報処理装置 10 のハードウェア構成例について説明する。図 16 は、本開示の一実施形態に係る情報処理装置 10 のハードウェア構成例を示すブロック図である。図 16 を参照すると、情報処理装置 10 は、例えば、CPU 871 と、ROM 872 と、RAM 873 と、ホストバス 874 と、ブリッジ 875 と、外部バス 876 と、インター

フェース 877 と、入力装置 878 と、出力装置 879 と、ストレージ 880 と、ドライブ 881 と、接続ポート 882 と、通信装置 883 と、を有する。なお、ここで示すハードウェア構成は一例であり、構成要素の一部が省略されてもよい。また、ここで示される構成要素以外の構成要素をさらに含んでもよい。

[0071] (CPU 871)

CPU 871 は、例えば、演算処理装置又は制御装置として機能し、ROM 872、RAM 873、ストレージ 880、又はリムーバブル記録媒体 901 に記録された各種プログラムに基づいて各構成要素の動作全般又はその一部を制御する。

[0072] (ROM 872、RAM 873)

ROM 872 は、CPU 871 に読み込まれるプログラムや演算に用いるデータ等を格納する手段である。RAM 873 には、例えば、CPU 871 に読み込まれるプログラムや、そのプログラムを実行する際に適宜変化する各種パラメータ等が一時的又は永続的に格納される。

[0073] (ホストバス 874、ブリッジ 875、外部バス 876、インターフェース 877)

CPU 871、ROM 872、RAM 873 は、例えば、高速なデータ伝送が可能なホストバス 874 を介して相互に接続される。一方、ホストバス 874 は、例えば、ブリッジ 875 を介して比較的データ伝送速度が低速な外部バス 876 に接続される。また、外部バス 876 は、インターフェース 877 を介して種々の構成要素と接続される。

[0074] (入力装置 878)

入力装置 878 には、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ、及びレバー等が用いられる。さらに、入力装置 878 としては、赤外線やその他の電波を利用して制御信号を送信することが可能なリモートコントローラ（以下、リモコン）が用いられることもある。また、入力装置 878 には、マイクロフォンなどの音声入力装置が含まれる。

[0075] (出力装置 879)

出力装置 879 は、例えば、CRT (Cathode Ray Tube)、LCD、又は有機 EL 等のディスプレイ装置、スピーカ、ヘッドホン等のオーディオ出力装置、プリンタ、携帯電話、又はファクシミリ等、取得した情報を利用者に対して視覚的又は聴覚的に通知することが可能な装置である。また、本開示に係る出力装置 879 は、触覚刺激を出力することが可能な種々の振動デバイスを含む。

[0076] (ストレージ 880)

ストレージ 880 は、各種のデータを格納するための装置である。ストレージ 880 としては、例えば、ハードディスクドライブ (HDD) 等の磁気記憶デバイス、半導体記憶デバイス、光記憶デバイス、又は光磁気記憶デバイス等が用いられる。

[0077] (ドライブ 881)

ドライブ 881 は、例えば、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリ等のリムーバブル記録媒体 901 に記録された情報を読み出し、又はリムーバブル記録媒体 901 に情報を書き込む装置である。

[0078] (リムーバブル記録媒体 901)

リムーバブル記録媒体 901 は、例えば、DVD メディア、Blu-ray (登録商標) メディア、HD DVD メディア、各種の半導体記憶メディア等である。もちろん、リムーバブル記録媒体 901 は、例えば、非接触型 IC チップを搭載した IC カード、又は電子機器等であってもよい。

[0079] (接続ポート 882)

接続ポート 882 は、例えば、USB (Universal Serial Bus) ポート、IEEE1394 ポート、SCSI (Small Computer System Interface)、RS-232C ポート、又は光オーディオ端子等のような外部接続機器 902 を接続するためのポートである。

[0080] (外部接続機器 902)

外部接続機器902は、例えば、プリンタ、携帯音楽プレーヤ、デジタルカメラ、デジタルビデオカメラ、又はICレコーダ等である。

[0081] (通信装置883)

通信装置883は、ネットワークに接続するための通信デバイスであり、例えば、有線又は無線LAN、Bluetooth(登録商標)、又はWUSB(Wireless USB)用の通信カード、光通信用のルータ、ADSL(Asymmetric Digital Subscriber Line)用のルータ、又は各種通信用のモデム等である。

[0082] <3. まとめ>

以上説明したように、本実施形態に係る情報処理装置は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された重みベクトルを用いて、当該重みベクトルと対応する入力ベクトルとの内積演算を行うことを特徴の一つとする。係る構成によれば、内積演算に係る処理負担をより軽減すると共に、重み係数の量子化粒度を保証することが可能となる。

[0083] 以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

[0084] また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

[0085] なお、以下のような構成も本開示の技術的範囲に属する。

(1)

複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく内積演算を行い出力値を算出する演算部、

を備え、

前記演算部は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出する、
情報処理装置。

(2)

前記演算部は、べき乗により表現された前記重み係数に基づいて前記出力値を算出する、
前記(1)に記載の情報処理装置。

(3)

前記重み係数の底は、前記演算部による内積演算において、0より大きく1未満の値に定まる、
前記(2)に記載の情報処理装置。

(4)

前記内積演算における乗算回数は、前記重み係数の底の値に基づいて定まる、
前記(2)または(3)に記載の情報処理装置。

(5)

前記演算部は、値の小さい順番に並び替えられた複数の前記重み係数に基づいて内積演算を行う、
前記(1)～(4)のいずれかに記載の情報処理装置。

(6)

前記演算部は、最も値の大きい前記重み係数で正規化された複数の前記重み係数に基づいて内積演算を行う、
前記(5)に記載の情報処理装置。

(7)

前記演算部は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持するテーブルに基づいて前記入力値を取得し内積演算を行う、

前記（５）または（６）に記載の情報処理装置。

（８）

前記アドレス情報は、アドレス間の相対位置を示すオフセットを含む、
前記（７）に記載の情報処理装置。

（９）

前記テーブルは、前記アドレス情報と対応付いた乗算指示情報をさらに保持する、

前記（７）または（８）に記載の情報処理装置。

（１０）

前記乗算指示情報は、少なくとも乗算の可否を指定する情報を含む、
前記（９）に記載の情報処理装置。

（１１）

前記テーブルは、前記アドレス情報と対応付いた符号情報をさらに保持する、

前記（７）～（１０）のいずれかに記載の情報処理装置。

（１２）

前記演算部は、ニューラルネットワークの順伝播に係る内積演算を行う、
前記（１）～（１１）のいずれかに記載の情報処理装置。

（１３）

複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づく
積和演算を実行する積和演算回路を備え、

前記積和演算回路は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持する記憶回路と、

前記アドレス情報により取得される前記入力値と前記重み係数に基づく演算を実行する乗算回路と、

を備える、

情報処理装置。

（１４）

プロセッサが、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づき積和演算を行い出力値を算出すること、

を含み、

前記算出することは、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出すること、

をさらに含む、

情報処理方法。

符号の説明

[0086]	1 0	情報処理装置
	1 1 0	入力部
	1 2 0	演算部
	1 3 0	記憶部
	1 4 0	出力部
	2 0 0	積和演算回路
	2 1 0	R A M
	2 2 0	加算回路
	2 3 0	アキュムレータ
	2 4 0	第1の乗算回路
	2 5 0	第2の乗算回路
	W T	アドレステーブル

請求の範囲

- [請求項1] 複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づき内積演算を行い出力値を算出する演算部、
を備え、
前記演算部は、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出する、
情報処理装置。
- [請求項2] 前記演算部は、べき乗により表現された前記重み係数に基づいて前記出力値を算出する、
請求項1に記載の情報処理装置。
- [請求項3] 前記重み係数の底は、前記演算部による内積演算において、0より大きく1未満の値に定まる、
請求項2に記載の情報処理装置。
- [請求項4] 前記内積演算における乗算回数は、前記重み係数の底の値に基づいて定まる、
請求項2に記載の情報処理装置。
- [請求項5] 前記演算部は、値の小さい順番に並び替えられた複数の前記重み係数に基づいて内積演算を行う、
請求項1に記載の情報処理装置。
- [請求項6] 前記演算部は、最も値の大きい前記重み係数で正規化された複数の前記重み係数に基づいて内積演算を行う、
請求項5に記載の情報処理装置。
- [請求項7] 前記演算部は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持するテーブルに基づいて前記入力値を取得し内積演算を行う、
請求項5に記載の情報処理装置。
- [請求項8] 前記アドレス情報は、アドレス間の相対位置を示すオフセットを含む、

請求項 7 に記載の情報処理装置。

[請求項9] 前記テーブルは、前記アドレス情報と対応付いた乗算指示情報をさらに保持する、

請求項 7 に記載の情報処理装置。

[請求項10] 前記乗算指示情報は、少なくとも乗算の可否を指定する情報を含む、

請求項 9 に記載の情報処理装置。

[請求項11] 前記テーブルは、前記アドレス情報と対応付いた符号情報をさらに保持する、

請求項 7 に記載の情報処理装置。

[請求項12] 前記演算部は、ニューラルネットワークの順伝播に係る内積演算を行う、

請求項 1 に記載の情報処理装置。

[請求項13] 複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づき積和演算を実行する積和演算回路を備え、

前記積和演算回路は、値の小さい順番に並び替えられた複数の前記重み係数と対応する前記入力値のアドレス情報を保持する記憶回路と

、

前記アドレス情報により取得される前記入力値と前記重み係数に基づき演算を実行する乗算回路と、

を備える、

情報処理装置。

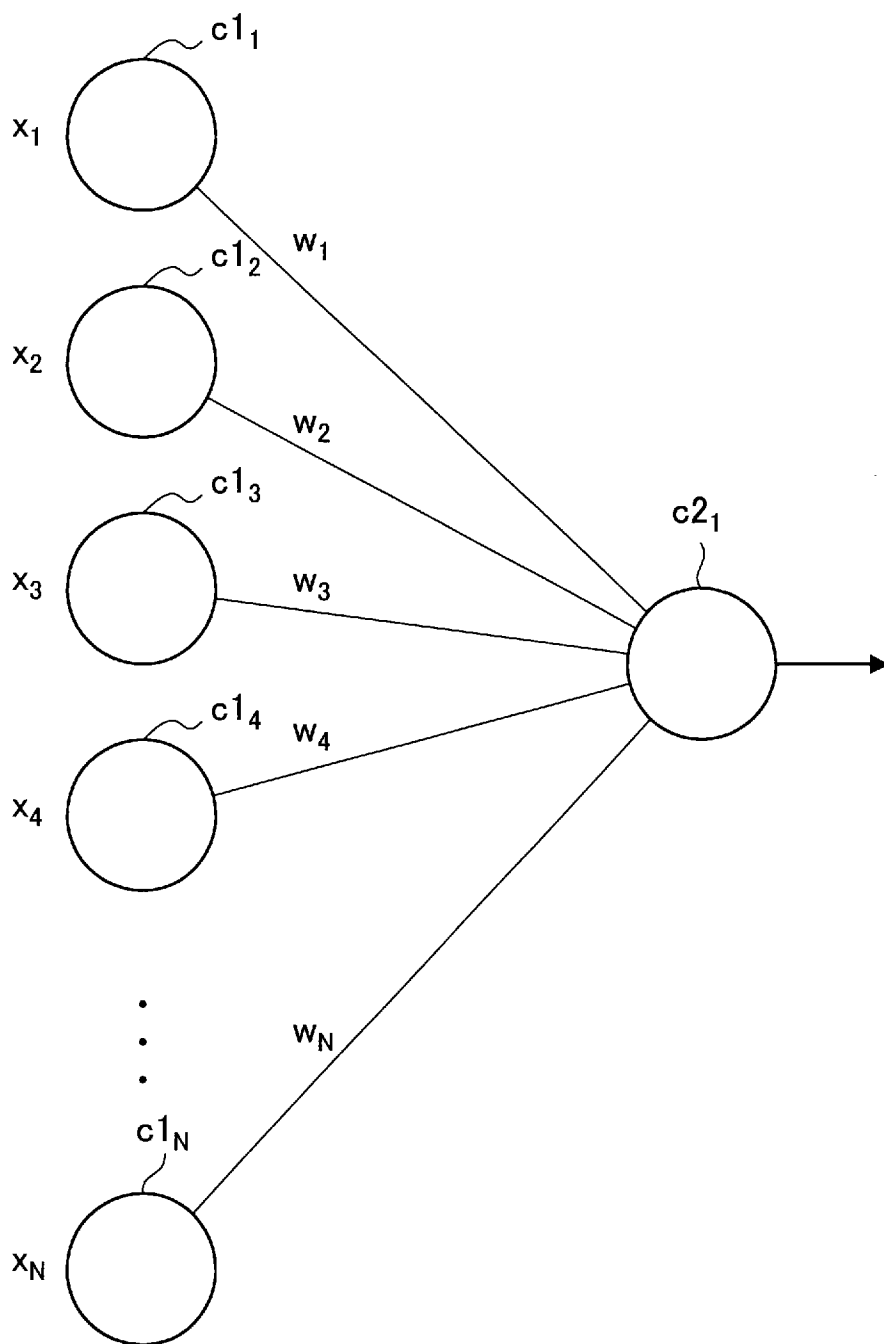
[請求項14] プロセッサが、複数の入力値と前記入力値にそれぞれ対応する複数の重み係数とに基づき積和演算を行い出力値を算出すること、

を含み、

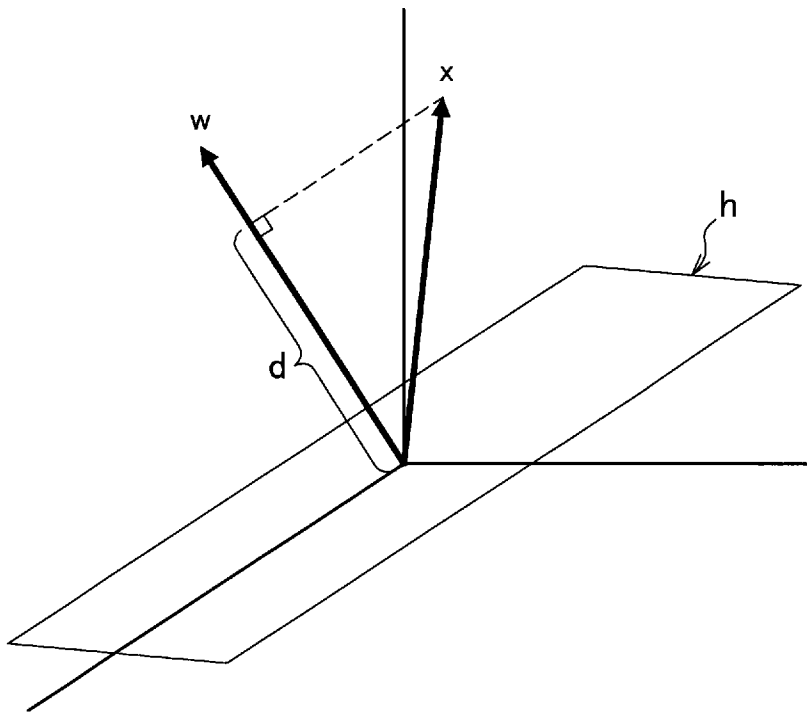
前記算出することは、N次元超球表面におけるベクトル方向の粒度に基づいて量子化された前記重み係数に基づいて前記出力値を算出すること、

をさらに含む、
情報処理方法。

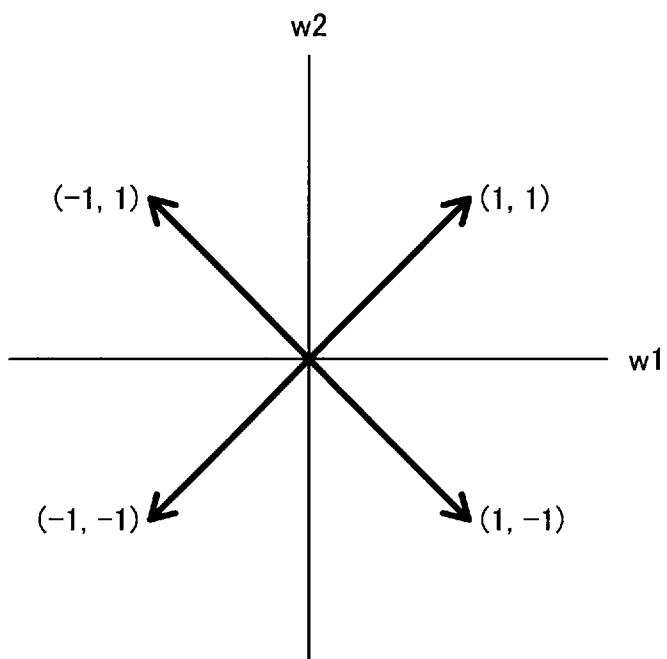
[図1]



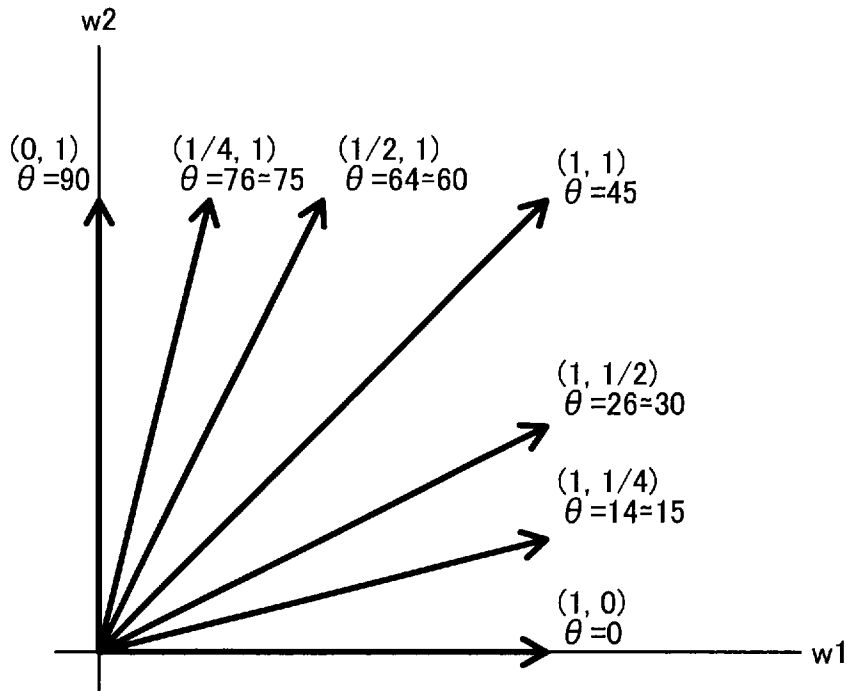
[図2]



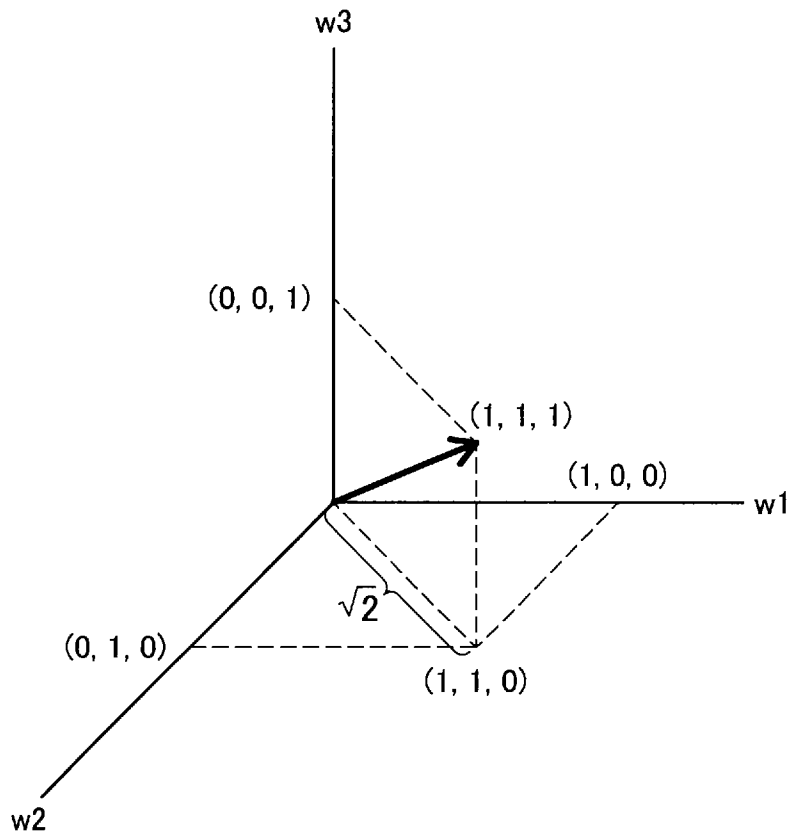
[図3]



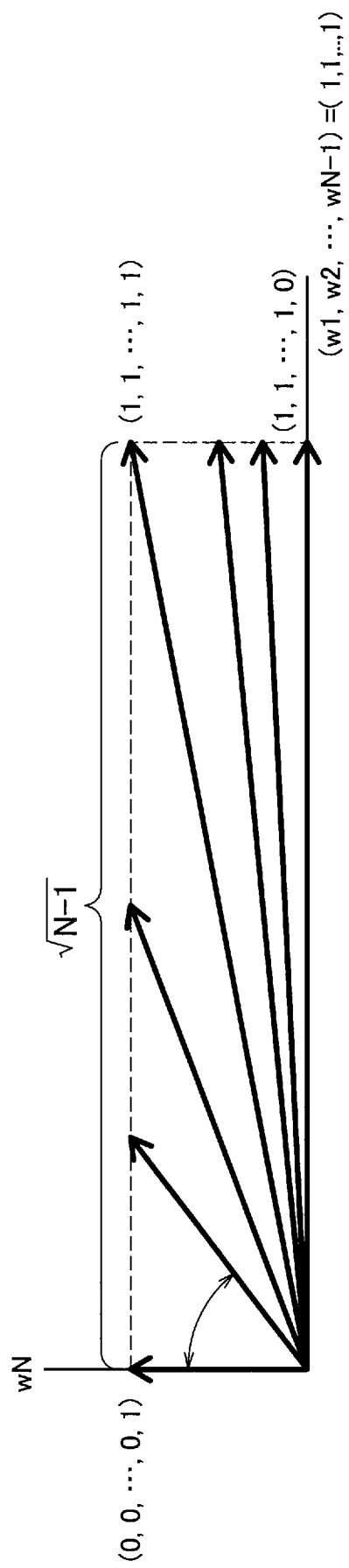
[図4]



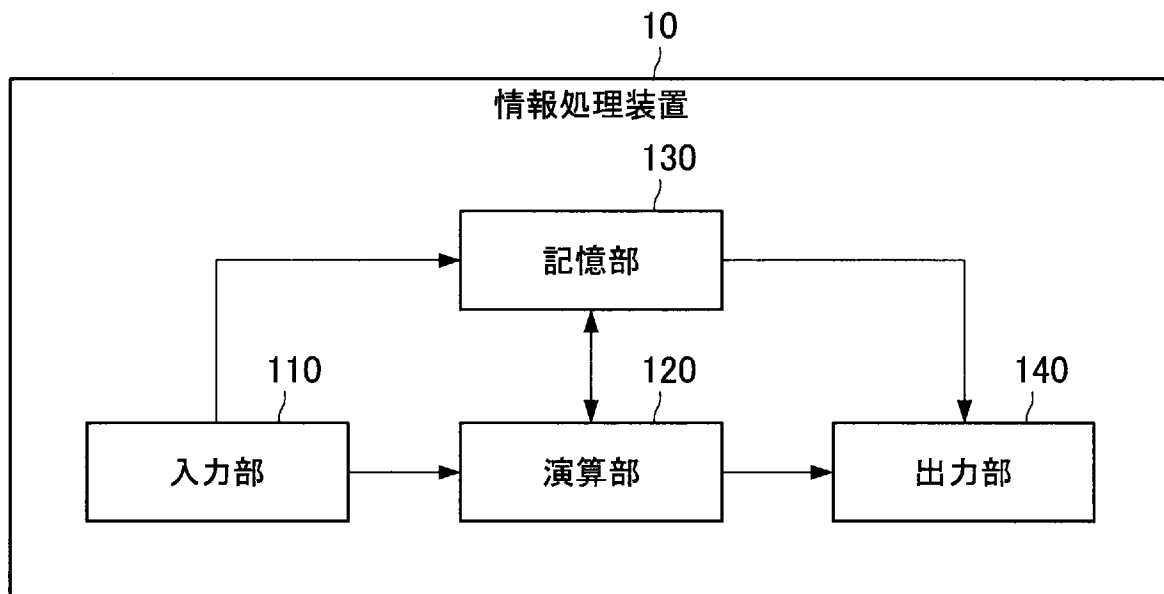
[図5]



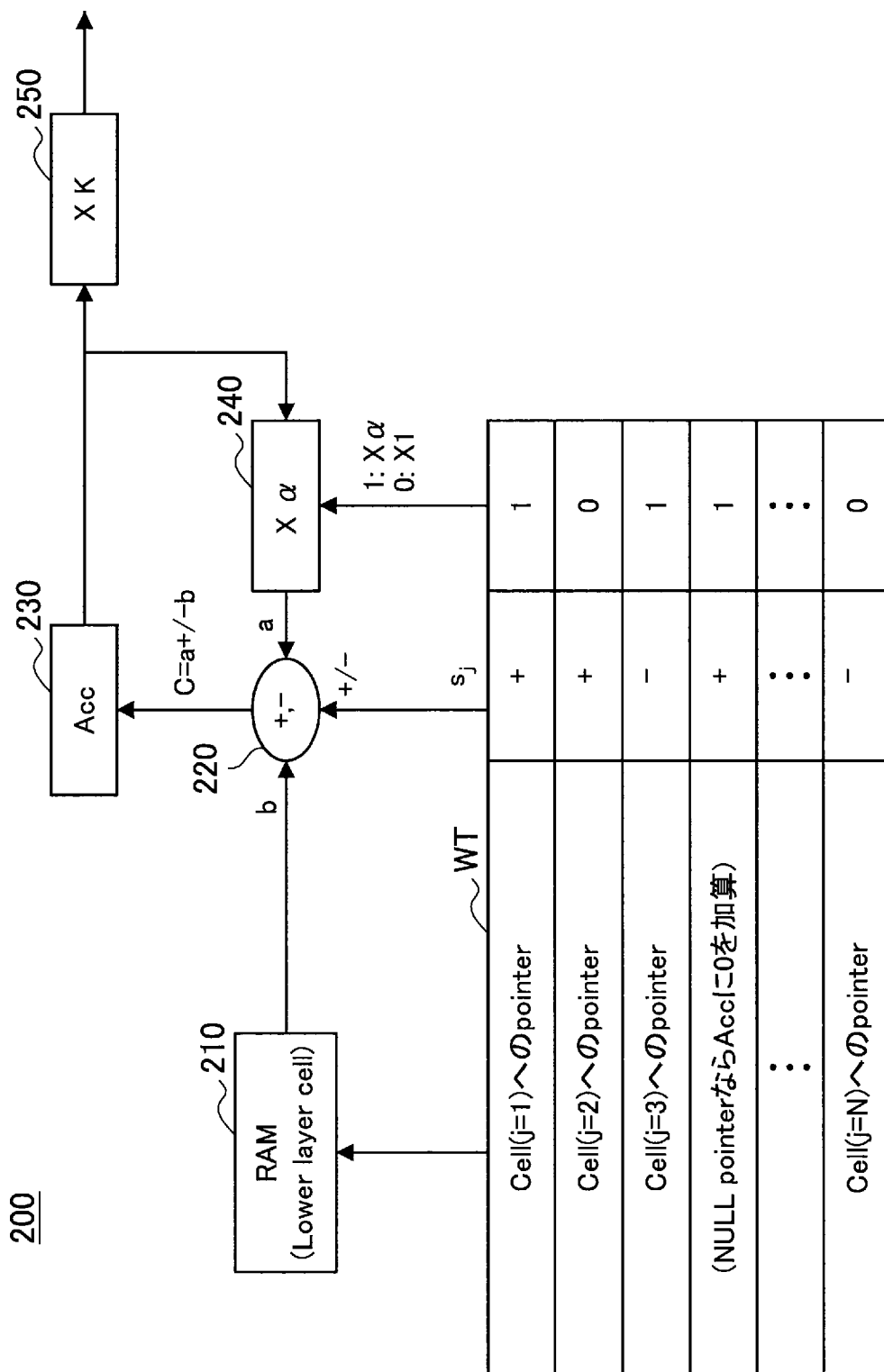
[図6]



[図7]



[図8]



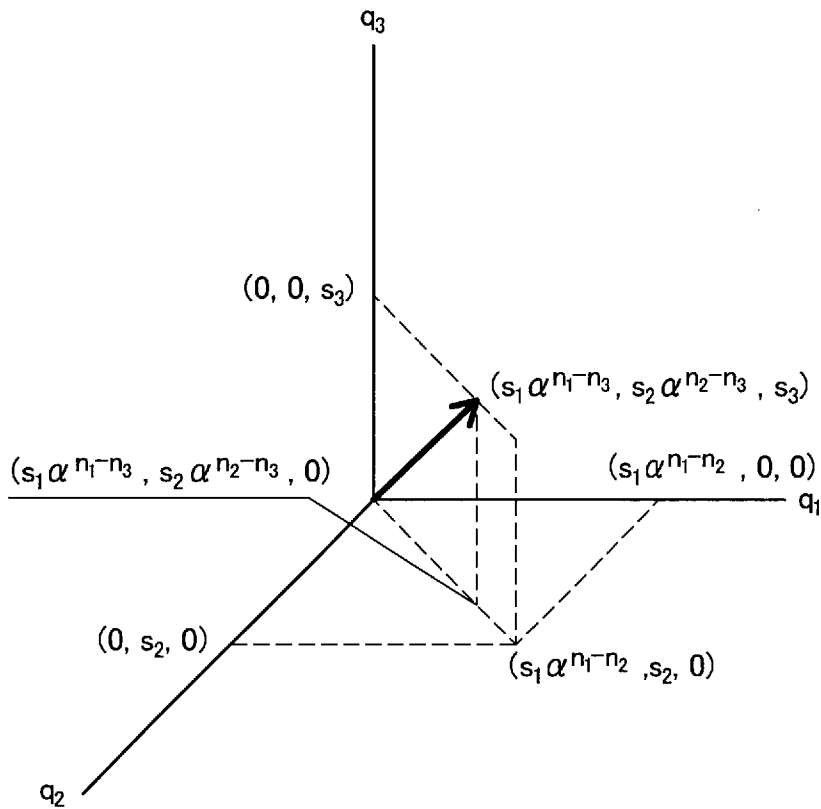
200

[図9]

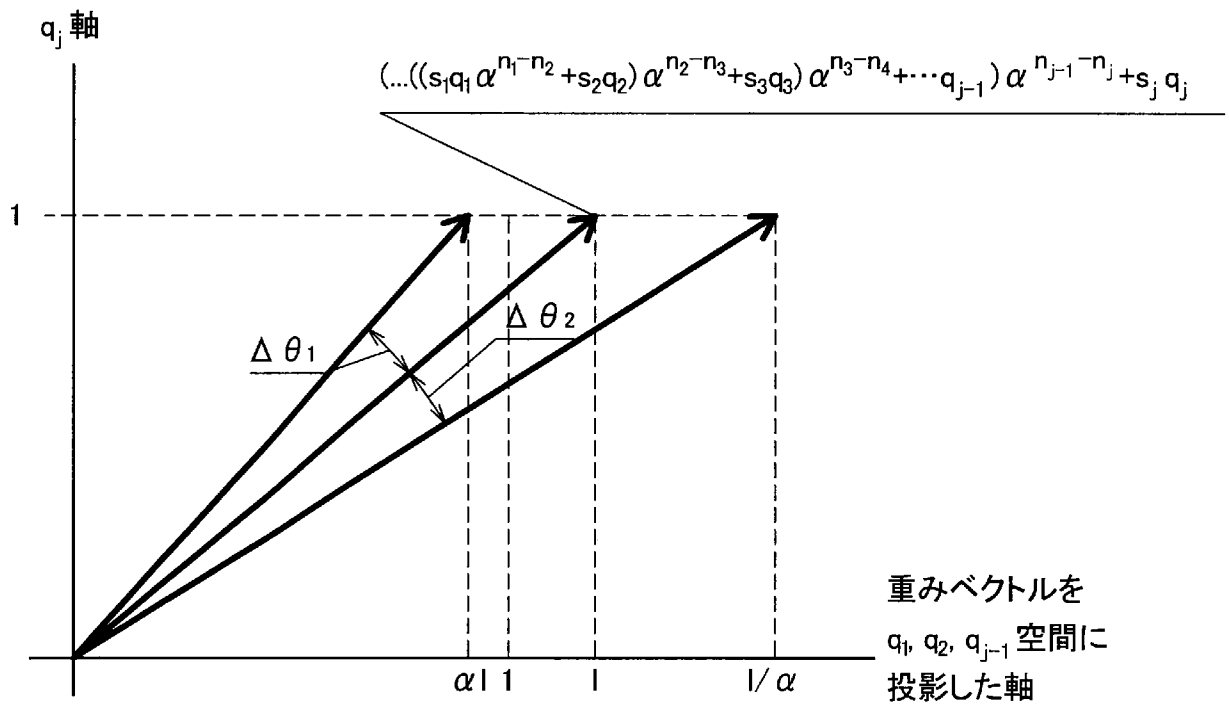
WT

Cell(j=1)へのpointer	+	1
Offset = 1	+	0
Offset = 2	-	0
Offset = 1	+	0
⋮	⋮	⋮
Cell(j=N)へのpointer	-	1

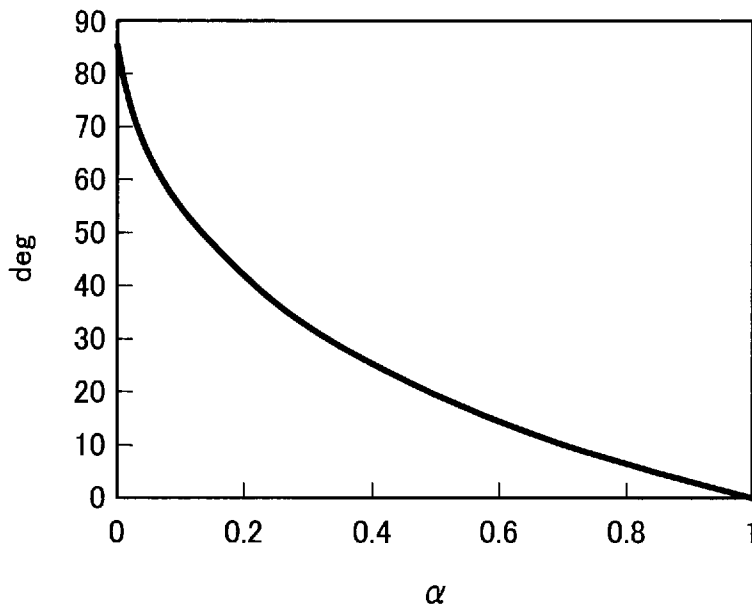
[図10]



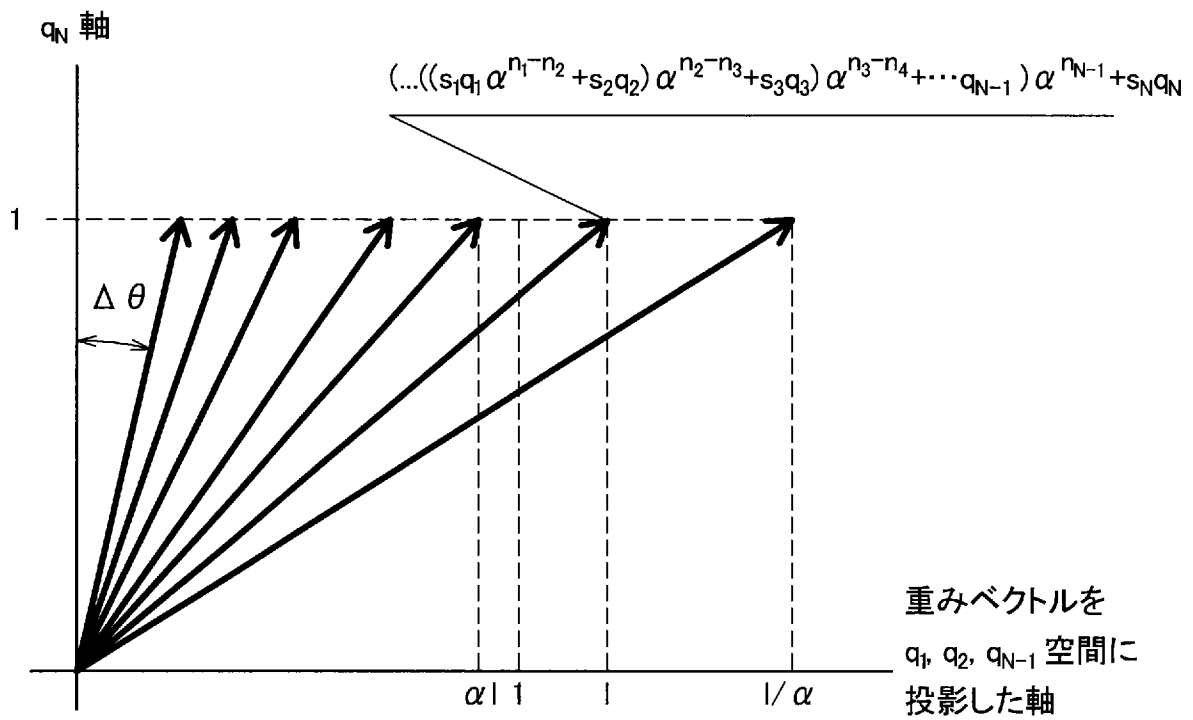
[図11]



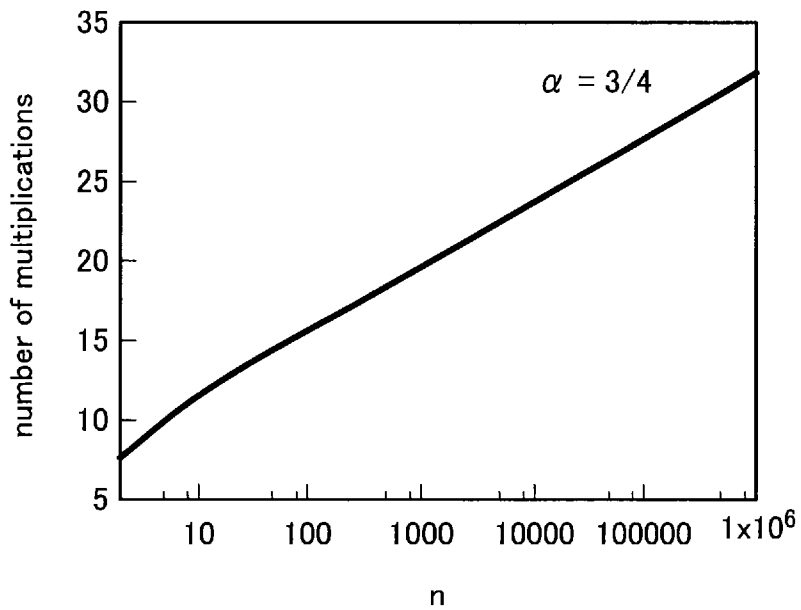
[図12]



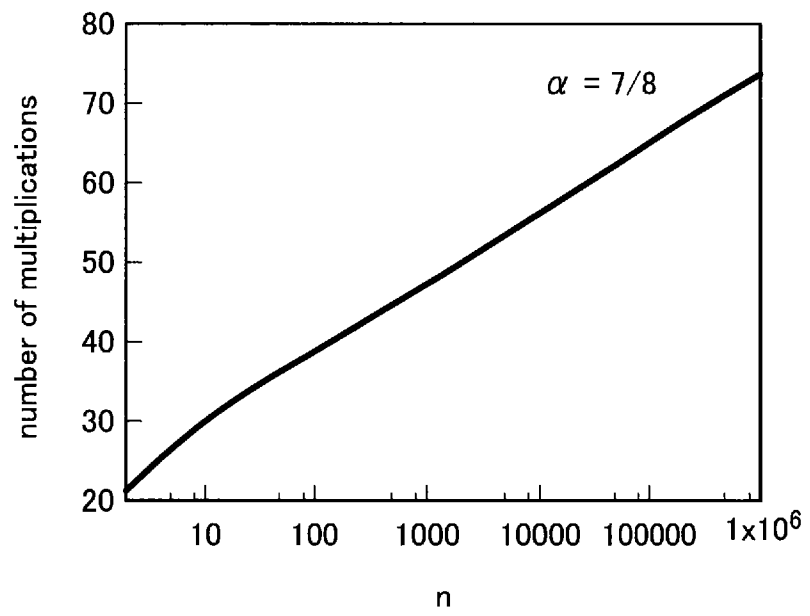
[図13]



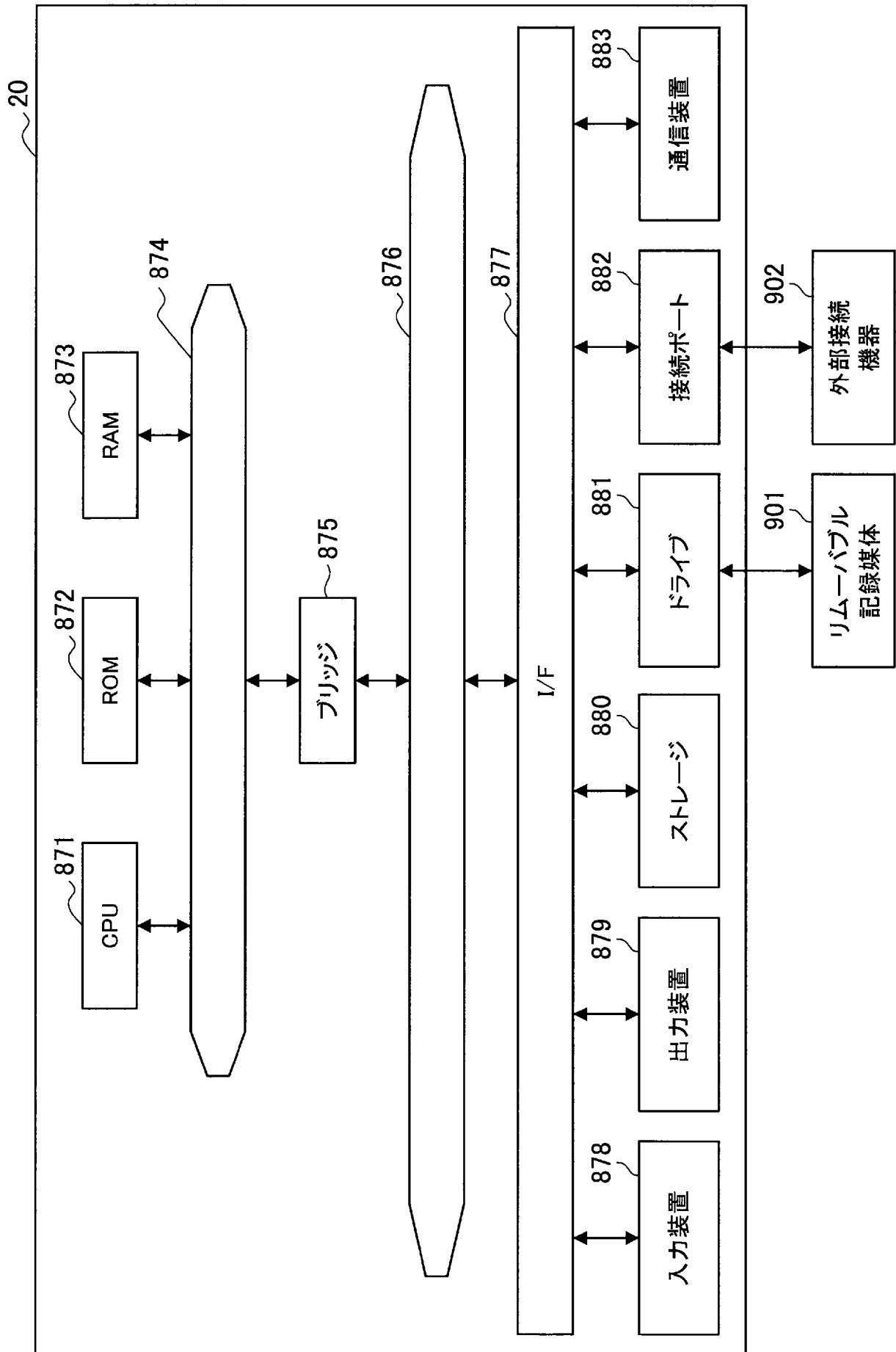
[図14]



[図15]



[図16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/002404

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06F17/10 (2006.01) i, G06N3/063 (2006.01) i, G06N99/00 (2010.01) 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. G06F17/10, G06N3/063, G06N99/00

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	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

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.
X A	MACHADO, Emerson et al., "FPGA Implementation of a Feedforward Neural Network-Based Classifier Using the xQuant Technique", Proceedings of 2017 IEEE 8th Latin American Symposium on Circuits and Systems, 20 February 2017, pages 1-4, Electronic ISSN: 2473-4667 in particular, abstract, page 2, right column, line 20 to page 3, left column, line 36	1-3, 12, 14 4-11
X A	JP 2006-180093 A (NEC ELECTRONICS CORP.) 06 July 2006, claims, paragraphs [0045], [0053], fig. 5, 8 & US 2006/0133303 A1 claims, paragraphs [0078], [0086], fig. 5, 8	13 5-11

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
06 April 2018 (06.04.2018)

Date of mailing of the international search report
17 April 2018 (17.04.2018)

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. G06F17/10(2006.01)i, G06N3/063(2006.01)i, G06N99/00(2010.01)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. G06F17/10, G06N3/063, G06N99/00

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	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018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
X A	MACHADO, Emerson et al., FPGA Implementation of a Feedforward Neural Network-Based Classifier Using the xQuant Technique, Proceedings of 2017 IEEE 8th Latin American Symposium on Circuits and Systems, 2017.02.20, pages1-4, Electronic ISSN: 2473-4667 特に、Abstract, 2ページ右欄20行~3ページ左欄36行	1-3, 12, 14 4-11
X A	JP 2006-180093 A (NECエレクトロニクス株式会社) 2006.07.06, 【特許請求の範囲】、【0045】、【0053】、【図5】、【図8】 & US 2006/0133303 A1 claims, [0078], [0086], FIG. 5, 8	13 5-11

☐ 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	の日の後に公表された文献
「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	「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
「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	「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
「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）	「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
「O」口頭による開示、使用、展示等に言及する文献	「&」同一パテントファミリー文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日
 06.04.2018

国際調査報告の発送日
 17.04.2018

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）	5B	9191
田中 幸雄		
電話番号 03-3581-1101 内線 3545		