

(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9月24日(24.09.2020)



(10) 国際公開番号

WO 2020/189704 A1

(51) 国際特許分類:
G06N 3/04 (2006.01)

(21) 国際出願番号: PCT/JP2020/011909

(22) 国際出願日: 2020年3月18日(18.03.2020)

(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 2019-052880 2019年3月20日(20.03.2019) 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). 国立大学法人東京大学 (THE UNIVERSITY OF TOKYO) [JP/JP]; 〒1138654 東京都文京区本郷七丁目3番1号 Tokyo (JP).

(72) 発明者: 酒見 悠介 (SAKEMI Yusuke); 〒1088001 東京都港区芝五丁目7番1号 日本電気株

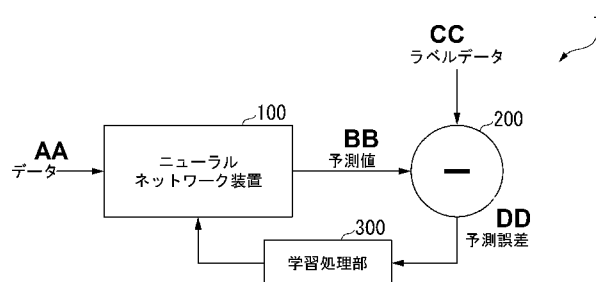
式会社内 Tokyo (JP). 森野 佳生 (MORINO Kai); 〒1138654 東京都文京区本郷七丁目3番1号 国立大学法人東京大学内 Tokyo (JP). 合原 一幸 (AIHARA Kazuyuki); 〒1138654 東京都文京区本郷七丁目3番1号 国立大学法人東京大学内 Tokyo (JP).

(74) 代理人: 棚井 澄雄, 外 (TANAI Sumio et al.); 〒1006620 東京都千代田区丸の内一丁目9番2号 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, 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,

(54) Title: NEURAL NETWORK DEVICE, NEURAL NETWORK SYSTEM, PROCESSING METHOD, AND RECORDING MEDIUM

(54) 発明の名称: ニューラルネットワーク装置、ニューラルネットワークシステム、処理方法および記録媒体



100 Neural network device
300 Learning processing unit
AA Data
BB Predicted value
CC Label data
DD Prediction error

(57) Abstract: This neural network device comprises: a neuron model unit configured as a leakless integrate-and-fire-type spiking neuron and a spiking neuron with which a postsynaptic current is represented using a step function, the neuron model unit being fired once at the most in one process of a neural network and indicating the output of the neural model unit itself at firing timing; and a transfer processing unit for transferring information between the neuron model units.

(57) 要約: ニューラルネットワーク装置が、漏れなし積分発火型のスパイクングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイクングニューロンとして構成され、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでニューロンモデル部自らの出力を示すニューロンモデル部と、前記ニューロンモデル部の間の情報伝達を行う伝達処理部と、を備える。

QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, 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))

明 細 書

発明の名称：

ニューラルネットワーク装置、ニューラルネットワークシステム、処理方法および記録媒体

技術分野

[0001] 本発明は、ニューラルネットワーク装置、ニューラルネットワークシステム、処理方法および記録媒体に関する。

背景技術

[0002] (順伝搬型スパイキングニューラルネットワークについて)

ニューラルネットワークの一形態として、順伝搬型スパイキングニューラルネットワーク (Spiking Neural Network ; S N N) がある。スパイキングニューラルネットワークとは、スパイキングニューロンモデル (スパイキングニューロン、または、単にニューロンとも称する) が結合しネットワークを形成したものである。

順伝搬型とは、ネットワーク構成方法の一つであり、層から層への結合における情報伝達が一方向のネットワークのことである。順伝搬型スパイキングニューラルネットワークの各層は1つ以上のスパイキングニューロンで構成されており、同層内のスパイキングニューロン間の結合は存在しない。

[0003] 図14は、順伝搬型スパイキングニューラルネットワークの階層構造の例を示す図である。図14は、順伝搬4層スパイキングニューラルネットワークの例を示している。但し、順伝搬型スパイキングニューラルネットワークの層数は、4層に限定されず2層以上であればよい。

図14に例示されるように順伝搬型スパイキングニューラルネットワークは階層構造に構成され、データの入力を受けて演算結果を出力する。ニューラルネットワークが出力する演算結果を予測値または予測とも称する。

ニューラルネットワークの第1層 (図14の例では、層1011) は入力層と呼ばれ、最後の層 (図14の例では、第4層 (層1014)) は出力層

と呼ばれる。入力層と出力層との間にある層（図14の例では、第2層（層1012）および第3層（層1013））は隠れ層と呼ばれる。

[0004] 図15は、順伝搬型スパイキングニューラルネットワークの構成例を示す図である。図15は、図14における4つの層（層1011～1014）が、それぞれ3つのスパイキングニューロン（スパイキングニューロンモデル）1021を有している場合の例を示している。但し、順伝搬型スパイキングニューラルネットワークが備えるスパイキングニューロンの個数は、特定の個数に限定されず、各層が1つ以上のスパイキングニューロンを備えていればよい。各層が同じ個数のスパイキングニューロンを備えていてもよいし、層によって異なる個数のスパイキングニューロンを備えていてもよい。

[0005] スパイキングニューロン1021は、生物学的神経細胞の細胞体部による信号の統合およびスパイクの生成（発火）を模擬する。

伝達経路1022は、生物学的神経細胞の軸索およびシナプスによる信号の伝達を模擬する。伝達経路1022は、隣り合う層間の2つのスパイキングニューロン1021を結んで配置され、前段層のスパイキングニューロン1021から後段層側のスパイキングニューロン1021へスパイクを伝達する。

[0006] 図15の例では、伝達経路1022は、層1011のスパイキングニューロン1021の各々から層1012のスパイキングニューロン1021の各々へ、層1012のスパイキングニューロン1021の各々から層1013のスパイキングニューロン1021の各々へ、および、層1013のスパイキングニューロン1021の各々から層1014のスパイキングニューロン1021の各々へ、スパイクを伝達する。

[0007] スパイキングニューロンモデルは、膜電位を内部状態として持ち、膜電位が微分方程式に従って時間発展するモデルである。一般的なスパイキングニューロンモデルとして、漏れ積分発火ニューロンモデルが知られており、式（1）のような微分方程式に従って時間発展する。

[0008]

[数1]

$$\frac{d}{dt}v_i^{(n)} = -\alpha_{leak}v_i^{(n)} + I_i^{(n)}, \quad I_i^{(n)} = \sum_j w_{ij}^{(n)}\kappa(t - t_j^{(n-1)})$$

・・・(1)

[0009] ここで、 $v_i^{(n)}$ は、第 n 層の i 番目のスパイキングニューロンモデルにおける膜電位を示す。 α_{leak} は、漏れ積分発火モデルにおける漏れの大きさを示す定数の係数である。 $I_i^{(n)}$ は、第 n 層の i 番目のスパイキングニューロンモデルにおけるシナプス後電流を示す。 $w_{ij}^{(n)}$ は、第 $n-1$ 層の j 番目のスパイキングニューロンモデルから第 n 層の i 番目のスパイキングニューロンモデルへの結合の強さを示す係数であり、重みと呼ばれる。

t は時刻を示す。 $t_j^{(n-1)}$ は第 $n-1$ 層の j 番目のニューロンの発火タイミング（発火時刻）を示す。 κ は前段の層から伝達されたスパイクがシナプス後電流へ与える影響を示す関数である。

[0010] 膜電位が閾値 V_{th} を超えると、そのスパイキングニューロンモデルはスパイクを生成し（発火）、その後、膜電位はリセット値 V_{reset} へと戻る。また、生成されたスパイクは、結合している後段層のスパイキングニューロンモデルへと伝達される。

[0011] 図16は、スパイキングニューロンの膜電位の時間発展の例を示す図である。図16のグラフの横軸は時刻を示し、縦軸は、膜電位を示す。図16は、第 n 層の i 番目のスパイキングニューロンの膜電位の時間発展の例を示しており、膜電位は、 $v_i^{(n)}$ と表されている。

上記のように、 V_{th} は、膜電位の閾値を示す。 V_{reset} は、膜電位のリセット値を示す。 $t_j^{(n-1)}$ は第 $n-1$ 層の j 番目のニューロンの発火タイミングを示す。 $t_1^{(n-1)}$ は第 $n-1$ 層の1番目のニューロンの発火タイミングを示す。 $t_2^{(n-1)}$ は第 $n-1$ 層の2番目のニューロンの発火タイミングを示す。 $t_3^{(n-1)}$ は第 $n-1$ 層の3番目のニューロンの発火タイミングを示す。

時刻 $t_1^{(n-1)}$ における1番目の発火および時刻 $t_3^{(n-1)}$ における3番目の発火では、何れも膜電位 $v_i^{(n)}$ は閾値 V_{th} に達していない。一方、時刻 t

$(n-1)_2$ における2番目の発火では、膜電位 $v^{(n)}_i$ が閾値 V_{th} に達し、その後すぐに、リセット値である V_{reset} に低下している。

[0012] スパイキングニューラルネットワークはCMOS (Complementary MOS) などハードウェア化したときに、深層学習モデルよりも消費電力を下げられると期待されている。その理由の一つは、人の脳は20ワット (W) 相当の低消費電力な計算媒体であり、スパイキングニューラルネットワークはそのような低消費電力の脳の活動を模倣することができるためである。

[0013] 脳相当の消費電力のハードウェアを作るには、脳の計算原理に倣い、スパイキングニューラルネットワークのアルゴリズムを開発していくことが必要である。例えば、画像認識を、スパイキングニューラルネットワークを用いて行えることが知られており、様々な教師あり学習アルゴリズムや教師なし学習アルゴリズムが開発されてきた。

[0014] (スパイキングニューラルネットワークにおける情報伝達方式について)

スパイキングニューラルネットワークのアルゴリズムでは、スパイクによる情報伝達方式において幾つかの手法があり、特に、頻度方式と時間方式とがよく用いられている。

頻度方式では、一定時間間隔の間に、ある特定のニューロンが何回発火したかで情報を伝達する。一方、時間方式では、スパイクのタイミングで情報を伝達する。

[0015] 図17は、頻度方式、時間方式それぞれにおけるスパイクの例を示す図である。図17の例で、頻度方式では、「1」、「3」、「5」の情報を、その情報に応じたスパイク数で示している。一方、時間方式では、「1」、「3」、「5」の情報の何れの場合もスパイク数は1つであり、情報に応じたタイミングでスパイクを生成することで、その情報を示している。図17の例では、ニューロンは、情報としての数が大きくなるほど遅いタイミングでスパイクを生成している。

[0016] 図17に示すように、時間方式は、頻度方式に比べて、少ないスパイク数で情報を表すことができる。非特許文献1では、画像認識等のタスクにおい

て、時間方式は頻度方式の10分の1以下のスパイク数で実行できることが報告されている。

ハードウェアの消費電力は、スパイク数増加によって増加するため、時間方式のアルゴリズムを用いると消費電力を削減することができる。

[0017] (順伝搬型スパイキングニューラルネットワークによる予測について)

順伝搬型スパイキングニューラルネットワークを用いることで、様々な課題を解くことができることが報告されている。例えば、図14のようなネットワーク構成において、入力層に画像データを入力し、スパイキングニューラルネットワークが、答えを予測するようにできる。時間方式の場合、予測値の出力方法として、例えば、出力層のニューロンのうち最も早く発火（スパイクを生成）したニューロンによって予測値を示すことができる。

[0018] (順伝搬型スパイキングニューラルネットワークの学習について)

スパイキングニューラルネットワークが正しく予測を行うには学習プロセスが必要である。例えば、画像を認識する学習プロセスでは、画像データと、その解答であるラベルデータとが用いられる。

[0019] 学習プロセスで、スパイキングニューラルネットワークは、データの入力を受けて予測値を出力する。そして、スパイキングニューラルネットワークに学習を行わせるための学習機構が、スパイキングニューラルネットワークが出力する予測値とラベルデータ（正解）との差である予測誤差を算出する。学習機構は、予測誤差から定義される損失関数 L を、スパイキングニューラルネットワークにおけるネットワークの重みの最適化によって最小化することで、スパイキングニューラルネットワークに学習を行わせる。

[0020] (損失関数の最小化について)

例えば、学習機構が、式(2)のように重みを更新することで、損失関数 L を最小化することができる。

[0021] [数2]

$$\Delta w_{ij}^{(n)} = -\eta \frac{\partial L}{\partial w_{ij}^{(n)}} \quad \dots (2)$$

[0022] ここで、 $\Delta w^{(n)}_{ij}$ は、重み $w^{(n)}_{ij}$ の増減を示す。 $\Delta w^{(n)}_{ij}$ の値が正の場合、重み $w^{(n)}_{ij}$ を増加させる。 $\Delta w^{(n)}_{ij}$ の値が負の場合、重み $w^{(n)}_{ij}$ を減少させる。

η は学習係数と呼ばれる定数である。

[0023] (確率的勾配降下法について)

確率的勾配降下法では、一部の訓練データを用いて重みを一度更新する。訓練データを全て用いて重みの更新を複数回繰り返したとき、その繰り返し単位をエポックと呼ぶ。確率的勾配降下法では、一般に、数十から数百のエポックを実行して学習を収束させる。また、一つのデータ（1つの入力データと1つのラベルデータ）で重みを更新することをオンライン学習と呼び、二つ以上のデータを用いて更新することをミニバッチ学習と呼ぶ。

[0024] (予測結果の出力について)

前述のように、順伝搬型スパイキングニューラルネットワークを用いることで、様々な課題を解くことができることが報告されている。例えば上記のように、入力層に画像データを入力し、ネットワークが、その画像の答えを予測するようにできる。

[0025] 図18は、スパイキングニューラルネットワークの予測結果の出力表現の例を示す図である。

例えば、0から2までの3個の数字の画像を認識するタスクにおいては、図18に示すように、3個のニューロンが出力層を構成し、それぞれが0から2までの数字に対応しており、そのうち最も早く発火したニューロンが示す数字がネットワークの示す予測となる。なお、このネットワークの動作は、ニューロンの発火タイミングによって情報がコーディングされているので、時間方式である。

[0026] (非線形関数とハードウェア実装について)

スパイキングニューラルネットワークの専用ハードウェアは、一般に、ニューロモルフィックハードウェアとよばれている。このハードウェアの実装について、アナログ回路による実装や、デジタル回路による実装が知られて

いる。

ハードウェアでは、一般に、その消費電力および回路面積を小さくすることが求められる。しかし、一方で、複雑なニューロンモデルや、複雑な学習則を実装すると、消費電力および回路面積が大きくなってしまう。

[0027] (非線形関数について)

ニューロンモデルでは、生物学的なニューロンとの整合性から、非線形関数を含む形が採用されることが多い。

[0028] (データの移動について)

ニューロモルフィックハードウェアの消費電力では、重みなどのメモリデータの移動が大きな寄与を占めている。そのため、学習則において、データの移動が少ないアルゴリズムを用いることで消費電力を下げられる。データの移動を少なくするには、移動の回数を少なくすること、および、データの移動距離を小さくすることのうち、何れか一方、もしくは両方を行えばよい。

[0029] 図19は、予測時および学習時におけるデータの移動の例を示す図である。図19では、ニューロンを三角で示し、重みを丸で示している。予測時は、実線で示すようなデータの移動が発生する。一方、学習時、特に、重み w_1 の更新時には、破線で示すようなデータの移動が発生する。

[0030] (リークなしモデルについて)

非特許文献2では、式(1)の定数 α_{leak} を0とおく漏れ(リーク)なし積分発火モデルを用いることで認識精度を向上させる報告があった。非特許文献2では、漏れなし積分発火モデルとして、式(3)に示されるモデルが用いられている。

[0031] [数3]

$$\frac{d}{dt}v_i^{(n)} = I_i^{(n)}(t), \quad I_i^{(n)}(t) = \sum_j w_{ij}^{(n)} \theta(t - t_j^{(n-1)}) \exp\left(-\frac{t - t_j^{(n-1)}}{\tau}\right) \dots (3)$$

[0032] ここで、 $\exp$ は、自然指数関数である。 τ は定数を示す。

先行技術文献

非特許文献

[0033] 非特許文献1：T. Liu、外5名、“MT-spike: A multilayer time-based spiking neuromorphic architecture with temporal error backpropagation”、Proceedings of the 36th International Conference on Computer-Aided Design、IEEE Press、2017年、p. 450-457

非特許文献2：H. Mostafa、“Supervised Learning Based on Temporal Coding in Spiking Neural Networks”、IEEE Transactions on Neural Networks and Learning Systems、第29、2018年、p. 3227-3235

非特許文献3：S. M. Bohte、外2名、“Error-backpropagation in temporally encoded networks of spiking neurons”、Neurocomputing、第48巻、2002年、p. 17-37

発明の概要

発明が解決しようとする課題

[0034] ニューラルネットワークのモデルを簡単なものにできることが好ましい。

例えば、非特許文献2における、上記の式(3)に示される漏れなし積分モデルが、非線形関数($\exp(-x/\tau)$)を含むのに対し、この非線形関数なしにモデルを構成できることがモデルの簡単化の観点から好ましい。

[0035] 本発明は、上述の課題を解決することのできるニューラルネットワーク装置、ニューラルネットワークシステム、ニューラルネットワーク処理方法および記録媒体を提供することを目的としている。

課題を解決するための手段

[0036] 本発明の第1の態様によれば、ニューラルネットワーク装置は、漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンとして構成され、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでニューロンモデル手段自らの出力を示すニューロンモデル手段と、前記ニューロン

モデル手段の間の情報伝達を行う伝達処理手段と、を備える。

[0037] 本発明の第2の態様によれば、処理方法は、漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンであって、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでスパイキングニューロン自らの出力を示すスパイキングニューロンの動作を実行する工程と、前記スパイキングニューロン間の情報伝達を行う工程と、を含む。

[0038] 本発明の第3の態様によれば、記録媒体は、ASICに、漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンであって、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでスパイキングニューロン自らの出力を示すスパイキングニューロンの動作を実行する工程と、前記スパイキングニューロン間の情報伝達を行う工程と、を実行させるためのプログラムを記憶する。

発明の効果

[0039] 本発明によれば、ニューラルネットワークのモデルを比較的簡単なものにすることができる。

図面の簡単な説明

[0040] [図1]実施形態に係るニューラルネットワーク装置の階層構造の例を示す図である。

[図2]実施形態に係るニューラルネットワーク装置の構成例を示す図である。

[図3]実施形態に係るニューラルネットワークシステムの概略構成の例を示す図である。

[図4]実施形態に係るスパイクのタイミングと発火確率密度との関係を示す図である。

[図5]実施形態に係る重みが増加したときの発火確率密度の変化を示す図である。

[図6]実施形態に係るスパイクのタイミングが増加したときの発火確率密度の

変化を示す図である。

[図7]実施形態に係るネットワークの重みの更新則の例を示す図である。

[図8]実施形態に係るニューラルネットワーク装置のシミュレーション結果の例を示す図である。

[図9]実施形態に係る重みの変化に伴って膜電位が変化する様子を表す図である。

[図10]実施形態に係る発火タイミングの変化に伴って膜電位が変化する様子を表す図である。

[図11]実施形態に係る実施形態に係るニューラルネットワーク装置の構成例を示す図である。

[図12]少なくとも1つの実施形態に係る専用ハードウェアの構成例を示す概略ブロック図である。

[図13]少なくとも1つの実施形態に係るコンピュータの構成例を示す概略ブロック図である。

[図14]順伝搬型スパイキングニューラルネットワークの階層構造の例を示す図である。

[図15]順伝搬型スパイキングニューラルネットワークの構成例を示す図である。

[図16]スパイキングニューロンの膜電位の時間発展の例を示す図である。

[図17]頻度方式、時間方式それぞれにおけるスパイクの例を示す図である。

[図18]スパイキングニューラルネットワークの予測結果の出力表現の例を示す図である。

[図19]予測時および学習時におけるデータの移動の例を示す図である。

発明を実施するための形態

[0041] 以下、本発明の実施形態を説明するが、以下の実施形態は請求の範囲にかかる発明を限定するものではない。また、実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

[0042] (実施形態に係るニューラルネットワーク装置の構造について)

図1は、実施形態に係るニューラルネットワーク装置の階層構造の例を示す図である。

図1の例で、ニューラルネットワーク装置100は順伝搬4層スパイキングニューラルネットワーク（Spiking Neural Network；SNN）に構成されている。但し、ニューラルネットワーク装置100の層数は、図1に示す4層に限らず2層以上であればよい。

[0043] 図1に示すニューラルネットワーク装置100は、順伝搬型スパイキングニューラルネットワークとして機能し、データの入力を受けて演算結果（予測値、または、予測と称する）を出力する。

ニューラルネットワーク装置100の各層のうち、第1層（層111）は入力層に該当する。最後の層（第4層、層114）は出力層に該当する。入力層と出力層との間にある層（第2層（層112）および第3層（層113））は隠れ層に該当する。

[0044] 図2は、ニューラルネットワーク装置100の構成例を示す図である。図2は、図1における4つの層（層111～114）が、それぞれ3つのノード（ニューロンモデル部121）を有している場合の例を示している。但し、ニューラルネットワーク装置100が備えるニューロンモデル部121の個数は、特定の個数に限定されず、各層が2つ以上のニューロンモデル部121を備えていればよい。各層が同じ個数のニューロンモデル部121を備えていてもよいし、層によって異なる個数のニューロンモデル部121を備えていてもよい。

[0045] ニューロンモデル部121は、スパイキングニューロン（スパイキングニューロンモデル）として構成され、細胞体部による信号の統合およびスパイクの生成（発火）を模擬する。

伝達処理部122は、軸索およびシナプスによる信号の伝達を模擬する。伝達処理部122は、任意の層間の2つのニューロンモデル部121を結んで配置され、前段層側のニューロンモデル部121から後段層側のニューロンモデル部121へスパイクを伝達する。

[0046] 図2の例では、伝達処理部122は、層111のニューロンモデル部121の各々から層112のニューロンモデル部121の各々へ、層112のニューロンモデル部121の各々から層113のニューロンモデル部121の各々へ、および、層113のニューロンモデル部121の各々から層114のニューロンモデル部121の各々へ、スパイクを伝達する。

[0047] (実施形態に係るニューラルネットワークシステムの構成について)

実施形態に係るニューラルネットワークシステムは、学習プロセスを実行するために、例えば図3に示す構成を有する。

図3は、実施形態に係るニューラルネットワークシステムの概略構成の例を示す図である。図3に示す構成で、ニューラルネットワークシステム1は、ニューラルネットワーク装置100と、予測誤差算出部200と、学習処理部300を備える。

[0048] かかる構成で、ニューラルネットワーク装置100は、データの入力を受けて予測値を出力する。予測誤差算出部200は、ニューラルネットワーク装置100が出力する予測値とラベルデータ（正解）との差である予測誤差を算出し、学習処理部300へ出力する。学習処理部300は、予測誤差から定義される損失関数 L を、ニューラルネットワーク装置100のネットワークの重みの最適化によって最小化することで、ニューラルネットワーク装置100に学習を行わせる。

ニューラルネットワーク装置100と、学習処理部300とは、別々の装置として構成されていてもよいし、1つの装置として構成されていてもよい。

[0049] (実施形態に係るニューロンのモデル)

実施形態に係るスパイキングニューロンのモデル（ニューロンモデル部121）の説明を行う。ニューロンモデル部121として、リークのないスパイキングニューロンのモデルを用いる。このモデルを式(4)のように定義する。

[0050]

[数4]

$$\frac{d}{dt}v_i^{(m)} = I_i^{(m)}, \quad I_i^{(m)}(t) = \sum_j w_{ij}^{(m)} \theta(t - t_j^{(m-1)}) \quad \dots (4)$$

[0051] ここで、 $v_i^{(m)}$ は、第 m 層の i 番目のニューロンモデル部 121 における膜電位を示す。

$I_i^{(m)}$ は、第 m 層の i 番目のニューロンモデル部 121 におけるシナプス後電流を示す。上記のように、 t は時刻を示す。 $I_i^{(m)}(t)$ は、シナプス後電流 $I_i^{(m)}$ を時刻 t の関数として表したものである。

$w_{ij}^{(m)}$ は、第 $m-1$ 層の j 番目のニューロンモデル部 121 から第 m 層の i 番目のニューロンモデル部 121 への結合の強さを示す係数（重み）である。
 $t_j^{(m-1)}$ は、第 $m-1$ 層の j 番目のニューロンモデル部 121 の発火タイミングを示す。 θ はステップ関数を示す。

[0052] ステップ関数 θ は、式 (5) のように示される。

[0053] [数5]

$$\theta(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \quad \dots (5)$$

[0054] ステップ関数 $\theta(t)$ は、 $t \geq 0$ の場合は $\theta(t) = 1$ との定数値となり、 $t < 0$ の場合は $\theta(t) = 0$ との定数値となる関数であり、例えば $\exp(-x/\tau)$ といった非線形関数と比較して簡単な処理で演算できる。

[0055] 上記のように、ニューラルネットワーク装置 100 のネットワークは、順伝搬型の多層ネットワークとして構成される。また、ニューロンモデル部 121 の各々は、ニューラルネットワーク装置 100 への 1 つの入力に対して、最大で一回しか発火しないものとする。

また、ニューラルネットワーク装置 100 の出力は、出力層のニューロンモデル部 121 の発火タイミングによって示されるものとする。例えば、ニューラルネットワーク装置 100 の出力が、図 18 を参照して説明した表現方法を用いて示されていてもよい。

[0056] （実施形態に係るニューロンのモデルの効果）

ニューロンモデル部 121 によれば、式（４）に示されるようにステップ関数の重み付き線形和で表される比較的簡単なモデルとすることができる。

例えば、式（４）に示されるモデルは、式（３）に示されるモデルよりも簡単であると評価できる。

[0057] ニューロンモデル部 121 の処理をソフトウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済む。また、ニューロンモデル部 121 の処理をハードウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済むことに加えて、ハードウェアの回路面積が比較的小さくて済む。

また、ニューロンモデル部 121 によれば、リークを含まないモデルである点で、認識精度が高い。

また、ニューロンモデル部 121 は、時間方式による点で、頻度方式による場合よりも消費電力が少なくて済む。

[0058] （実施形態に係る出力層学習（１））

次に、ニューラルネットワークシステム 1 における学習アルゴリズムの説明を行う。

（Spike Prop について）

前述の式（２）の重み更新則中の導関数 $\partial L / \partial w^{(n)}_{ij}$ を導く手法として Spike Prop アルゴリズムが知られている（非特許文献 3 参照）。例えば、損失関数 L を、最終層のニューロンの発火タイミングを用いて式（６）のように定義する。

[0059] [数 6]

$$L = \sum_i L_i, \quad L_i = \frac{1}{2} \left(t_i^{(N)} - t_i^{(t)} \right)^2 \quad \dots (6)$$

[0060] ここで、 $t^{(N)}_i$ は、出力層の i 番目のニューロンの発火タイミングを示す

。なお、出力層を第N層として「N」で表記する。

$t^{(t)}_i$ は、i 番目の教師信号の発火タイミング（教師信号における、出力層の i 番目のニューロンの発火タイミング）を示す。また、ここでは、式（3）に示されるリークなしのニューロンモデルを対象にする。

損失関数の重みによる微分はチェインルールを用いて、式（7）のように示される。

[0061] [数7]

$$\frac{\partial L}{\partial w_{ij}^{(n)}} = \frac{\partial t_i^{(n)}}{\partial w_{ij}^{(n)}} \frac{\partial L}{\partial t_i^{(n)}} = \frac{\partial t_i^{(n)}}{\partial w_{ij}^{(n)}} \delta_i^{(n)} \quad \dots (7)$$

[0062] ここでいう損失関数の重みによる微分は、損失関数を重みで微分したものである。

ここで、伝搬誤差を式（8）のように定義した。

[0063] [数8]

$$\delta_j^{(n)} = \frac{\partial L}{\partial t_j^{(n)}} = \begin{cases} (t_j^{(N)} - t_j^{(T)}), & \text{for } n = N \\ \sum_i \frac{\partial t_i^{(n+1)}}{\partial t_j^{(n)}} \delta_i^{(n+1)}, & \text{for } n = 1, 2, \dots, N-1 \end{cases} \quad \dots (8)$$

[0064] 結局、導関数を導くためには、 $\partial t^{(n)}_i / \partial w^{(n)}_{ij}$ と、 $\partial t^{(n+1)}_i / \partial t^{(n)}_j$ とを計算すればよい。Spike Prop法を用いて、 $\partial t^{(n)}_i / \partial w^{(n)}_{ij}$ は式（9）のように導出できる。

[0065]

[数9]

$$\begin{aligned}
\frac{\partial t_i^{(n)}}{\partial w_{ij}^{(n)}} &= \frac{\partial v_i^{(n)}}{\partial w_{ij}^{(n)}} \frac{\partial t_i^{(n)}}{\partial v_i^{(n)}} \\
&= -\left(t_i^{(n)} - t_j^{(n-1)}\right) \theta\left(t_i^{(n)} - t_j^{(n-1)}\right) \left(\frac{\partial v_i^{(n)}}{\partial t} \Big|_{t=t_i^{(n)}}\right)^{-1} \\
&= \frac{-\left(t_i^{(n)} - t_j^{(n-1)}\right) \theta\left(t_i^{(n)} - t_j^{(n-1)}\right)}{\sum_j w_{ij}^{(n)} \theta\left(t_i^{(n)} - t_j^{(n-1)}\right)} \\
&\quad \dots (9)
\end{aligned}$$

[0066] また、 $\partial t^{(n+1)}_i / \partial t^{(n)}_j$ は式 (10) のように導出できる。

[0067] [数10]

$$\begin{aligned}
\frac{\partial t_i^{(n+1)}}{\partial t_j^{(n)}} &= \frac{\partial v_i^{(n+1)}}{\partial t_j^{(n)}} \frac{\partial t_i^{(n+1)}}{\partial v_i^{(n+1)}} = \frac{w_{ij}^{(n+1)} \theta\left(t_i^{(n+1)} - t_j^{(n)}\right)}{\sum_j w_{ij}^{(n+1)} \theta\left(t_i^{(n+1)} - t_j^{(n)}\right)} \\
&\quad \dots (10)
\end{aligned}$$

[0068] 式 (9) および式 (10) に示すように Spike Prop アルゴリズムでは、 $\partial t^{(n)}_i / \partial w^{(n)}_{ij}$ および $\partial t^{(n+1)}_i / \partial t^{(n)}_j$ を計算するために、同層内の重みの総和を計算する必要がある。

これに対し、ニューラルネットワーク装置 100 では、 $\partial t^{(n)}_i / \partial w^{(n)}_{ij}$ および $\partial t^{(n+1)}_i / \partial t^{(n)}_j$ を近似することにより簡易化された学習則を用いる。この学習則の導出について説明する。

[0069] まず、第 n 層の i 番目のニューロン（ニューロンモデル部 121）の発火タイミングが、発火確率密度 $R^n_i(t)$ によって確率的に決まるとの仮定をおく。上記のように、 t は時刻を示す。観測された第 n 層の発火タイミング $t^{(n)}_i$ と、その前の層のニューロン（ニューロンモデル部 121）の発火タイミング $t^{(n-1)}_j$ とから、発火確率密度 $R^{(n)}_i(t)$ の関数形を推定する。

[0070] 本モデルでは、各ニューロン（ニューロンモデル部 121）は一回以下しか発火しないため、情報を持つのは第一発火タイミングである。そのため、

推定された発火関数（発火確率密度の関数形）から得られるニューロンの第一発火タイミング（第一発火時刻）の分布が最大値をとる時刻を、第 n 層の発火タイミング $t^{(n)}_i$ とする。

上記のモデルを仮定することで、重み $w^{(n)}_{ij}$ が変化したときの、発火確率密度の関数形の変化 $\delta R^{(n)}_i(t)$ が得られる。式（11）に示すように、この発火確率密度関数の変化から、発火タイミングの変化 $\delta t^{(n)}_i$ が得られる。

[0071] [数11]

$$\delta w^{(n)}_{ij} \rightarrow \delta R^{(n)}_i(t) \rightarrow \delta t^{(n)}_i \quad \dots (11)$$

[0072] 式（11）は、重み $w^{(n)}_{ij}$ の変化分に応じて発火確率密度 $R^{(n)}_i(t)$ の変化分が求まり、発火確率密度 $R^{(n)}_i(t)$ の変化分に応じて第 n 層の発火タイミング $t^{(n)}_i$ の変化分が求まるという関係を示している。この関係により、重み $w^{(n)}_{ij}$ の変化分から第 n 層の発火タイミング $t^{(n)}_i$ の変化分を求めることができる。

式（11）の関係から、式（12）のように偏微分の近似が得られる。

[0073] [数12]

$$\frac{\partial t^{(n)}_i}{\partial w^{(n)}_{ij}} \approx \frac{\delta t^{(n)}_i}{\delta w^{(n)}_{ij}} \quad \dots (12)$$

[0074] （実施形態に係る出力層学習（1）の例）

発火確率密度 $R^{(n)}_i(t)$ は、漏れなしのスパイキングニューロンモデルにおいて、膜電位の傾き（時間微分）で近似することができる。この近似は、式（13）のように示される。

[0075] [数13]

$$R^{(n)}_i(t) \approx \frac{dv^{(n)}_i}{dt} = I^{(n)}_i(t) = \sum_{j'} w^{(n)}_{ij'} \theta(t - t^{(n-1)}_{j'}) \quad \dots (13)$$

[0076] さらに、この関数を区分線形関数 $R_{linear}(t)$ に近似し、式（14）を

得る。

[0077] [数14]

$$R_i^{(n)}(t) \approx \sum_{j'} w_{ij'}^{(n)} \theta(t - t_{j'}^{(n)}) \approx \alpha(t - t') \theta(t - t') = R_{linear}(t) \quad \dots (14)$$

[0078] ここで、 α 、 t' は、何れも定数であり、図4に示すように、 $t' < t^{(n-1)}_j$ を満たすものとする。

図4は、スパイクのタイミングと発火確率密度との関係を示す図である。

図4の、上段は、 $t^{(n-1)}_j$ と $t^{(n)}_i$ とのタイミングを示している。中段は、推定される n 層の i 番目のニューロンの発火確率密度 $R_{linear}(t)$ を示している。下段は、推定される発火確率密度 $R_{linear}(t)$ から計算される第一発火のタイミングの確率分布を示している。図4の上段、中段、下段の何れも横軸は時刻を示す。中段、下段それぞれの縦軸は発火確率密度を示す。

発火確率密度が区分線形関数 $R_{linear}(t)$ で与えられたときの第一発火タイミングの確率は以下のように計算できる。すなわち、時刻 t までに一度も発火しない確率を $x(t)$ とおくと、これは、式(15)の微分方程式を満たす。

[0079] [数15]

$$\frac{dx}{dt} = -xR_{linear}(t) = \alpha x(t - t') \theta(t - t') \quad \dots (15)$$

[0080] 式(15)の微分方程式を解くと、式(16)のようになる。

[0081] [数16]

$$x = e^{-\frac{1}{2}\alpha(t-t')^2\theta(t-t')} \quad \dots (16)$$

[0082] よって、第一スパイク発火確率密度 $P_f(t)$ は、式(17)のように求めることができる。

[0083]

[数17]

$$P_f(t) = -\frac{dx}{dt} = \alpha(t-t')\theta(t-t')e^{-\frac{1}{2}\alpha(t-t')^2\theta(t-t')} \quad \dots (17)$$

[0084] なお、第一スパイク発火確率密度 $P_f(t)$ は非負であり、式 (18) のように確率の定義を満たしていることが分かる。

[0085] [数18]

$$\int_{-\infty}^{\infty} P_f(t) dt = -\int_{-\infty}^{\infty} \frac{dx}{dt} dt = [x]_{-\infty}^{\infty} = 1 - 0 = 1 \quad \dots (18)$$

[0086] 第一スパイク発火確率密度 $P_f(t)$ が最大値をとる時刻 t^* は、 $P_f(t)$ の時間微分が 0 ($\partial P_f(t) / \partial t = 0$) より、式 (19) のように示される。

[0087] [数19]

$$t^* = t' + \frac{1}{\alpha} \quad \dots (19)$$

[0088] この時刻 t^* が出力スパイク時刻と一致する条件を課すことで、式 (20) が得られる。

[0089] [数20]

$$\alpha = \frac{1}{t_i^{(n)} - t'} \quad \dots (20)$$

[0090] 次に、重みが変化したときのニューロン（ニューロンモデル部 121）の発火確率密度 $R_i^{(n)}(t)$ の変化は、式 (21) のように示される。

[0091] [数21]

$$\frac{\delta R_i^{(n)}}{\delta w_{ij}^{(n)}} \approx \frac{\delta \left(I_i^{(n)}(t) \right)}{\delta w_{ij}^{(n)}} = \frac{\partial I_i^{(n)}}{\partial w_{ij}^{(n)}} = \theta(t - t_j^{(n-1)}) \quad \dots (21)$$

[0092] 発火確率密度を区分線形関数 $R_{linear}(t)$ で表記して、その変化 δR_i

(t) は、式 (22) のように示される。

[0093] [数22]

$$R_{linear}(t) + \delta R_i(t) \approx \alpha(t - t')\theta(t - t') + \delta w_{ij}\theta(t - t_j^{n-1}) \dots (22)$$

[0094] この変化は、図5のように示される。

図5は、重みが変化したときの発火確率の変化を示す図である。具体的には、図5は、重み $W^{(n)}_{ij}$ が $\delta W^{(n)}_{ij}$ だけ変化したときの、発火確率密度 $R_{linear}(t)$ の変化を示している。

図5の横軸は時刻を示し、縦軸は発火確率密度を示す。線L11は重み $W^{(n)}_{ij}$ が変化する前の発火確率密度を示しており、線L12は重み $W^{(n)}_{ij}$ が変化した後の発火確率密度を示している。

この発火関数が与えられたときの、発火タイミングは、 $t^{(n)}_i + \delta t^{(n)}_i$ と表される。 $t^{(n)}_i + \delta t^{(n)}_i$ を求めるために解くべき方程式は、式(23)のように示される。

[0095] [数23]

$$\frac{dx}{dt} = \alpha x((t - t')\theta(t - t') + \frac{\delta w_{ij}}{\alpha}\theta(t - t_j^{n-1})) \dots (23)$$

[0096] あるいは、 $t^{(n)}_i + \delta t^{(n)}_i$ を求めるために解くべき方程式は、式(24)のように示される。

[0097] [数24]

$$\frac{dx}{dt} = \begin{cases} 0 & (t \leq t') \\ \alpha x(t - t') & (t' < t \leq t_j^{(n-1)}) \\ \alpha x\left(t - \left(t' - \frac{\delta w_{ij}}{\alpha}\right)\right) & (t_j^{(n-1)} < t) \end{cases} \dots (24)$$

[0098] 式(24)の解は、初期条件を $x(0) = 1$ として、式(25)のようになる。

[0099]

[数25]

$$x(t) = \begin{cases} 1 & (t \leq t') \\ e^{-\frac{\alpha}{2}(t-t')^2} & (t' < t \leq t_j^{(n-1)}) \\ Ae^{-\frac{\alpha}{2}\left(t - \left(t' - \frac{\delta w_{ij}}{\alpha}\right)\right)^2} & (t_j^{(n-1)} \leq t) \end{cases} \quad \dots (25)$$

[0100] 式(25)のAは、式(26)のように示される。

[0101] [数26]

$$A = \frac{e^{-\frac{\alpha}{2}(t^{n-1}-t')^2}}{e^{-\frac{\alpha}{2}\left(t^{n-1} - \left(t' - \frac{\Delta w}{\alpha}\right)\right)^2}} \quad \dots (26)$$

[0102] このとき第一スパイク発火確率密度 $P_f(t)$ が最大値をとる時刻 t^* は、式(27)のようになる。

[0103] [数27]

$$t^* = \begin{cases} t_i^{(n)} & \text{for } t_i^{(n)} \leq t_j^{(n-1)} \\ t_i^{(n)} - \frac{\delta w_{ij}^{(n)}}{\alpha} & \text{for } t_j^{(n-1)} < t_i^{(n)} \end{cases} \quad \dots (27)$$

[0104] 重みが $\delta w^{(n)}_{ij}$ 変化したとき推定される出力スパイクの時刻の変化は、式(28)のように示される。

[0105] [数28]

$$\delta t_i^{(n)} = -\frac{\delta w_{ij}^{(n)}}{\alpha} \theta(t_i^{(n)} - t_j^{(n-1)}) \quad \dots (28)$$

[0106] 偏微分の近似値として式(29)が得られる。

[0107] [数29]

$$\frac{\partial t_i^{(n)}}{\partial w_{ij}^{(n)}} \approx \frac{\delta t_i^{(n)}}{\delta w_{ij}^{(n)}} = -\frac{1}{\alpha} \theta(t_i^{(n)} - t_j^{(n-1)}) \quad \dots (29)$$

[0108] 次に、 $\partial t^{(n)}_i / \partial t^{(n-1)}_j$ の近似を行う。

上述したのと同様、式（30）に示すように、発火確率密度 R の変化を経由することで、 $\delta t^{(n-1)}_j$ と $t^{(n)}_i$ との関係を導くことで偏微分を近似する。

[0109] [数30]

$$\delta t^{(n-1)}_j \rightarrow \delta R \rightarrow t^{(n)}_i \quad \dots (30)$$

[0110] 図6は、スパイクのタイミングが変化したときの発火確率密度の変化を示す図である。

具体的には、図6は、前段層のニューロン（ニューロンモデル部121）の発火時間 $t^{(n-1)}_j$ が $\delta t^{(n-1)}_j$ だけ変化したとき、後段層のニューロン（ニューロンモデル部121）の発火確率密度 $R_{linear}(t)$ が変化する様子を表している。

図6の横軸は時刻を示し、縦軸は発火確率密度を示す。線L21は、変化する前の発火確率密度 $R_{linear}(t)$ を示しており、線L21は、変化後の発火確率密度 $R_{linear}(t) + \delta R_{linear}(t)$ を示している。

[0111] 発火確率密度を線形近似した区分線形関数 $R_{linear}(t)$ は、 $n-1$ 層の全てのニューロン（ニューロンモデル部121）から n 層の i 番目のニューロン（ニューロンモデル部121）へのスパイクを平均化したものであり、式（31）のように変形できる。

[0112] [数31]

$$\begin{aligned} R_{linear} &= \alpha(t - t')\theta(t - t') \\ &= \alpha \left(\frac{w_{ij}^{(n)}\theta(t - t_j^{(n-1)})}{\sum_{(i,j')} w_{ij'}^{(n)}\theta(t - t_{j'}^{(n-1)})} + \frac{\sum_{j' \neq j} w_{ij'}^{(n)}\theta(t - t_{j'}^{(n-1)})}{\sum_{(i,j')} w_{ij'}^{(n)}\theta(t - t_{j'}^{(n-1)})} \right) (t - t')\theta(t - t') \\ &= \left(\frac{w_{ij}^{(n)}\theta(t - t_j^{(n-1)})}{t_j^{(n-1)} - t'} + \frac{\sum_{j' \neq j} w_{ij'}^{(n)}\theta(t - t_{j'}^{(n-1)})}{t_j^{(n-1)} - t'} \right) (t - t')\theta(t - t') , \\ \alpha &= \frac{\sum_{(i,j')} w_{ij'}^{(n)}\theta(t - t_j^{(n-1)})}{t^{(n-1)} - t'} \end{aligned} \quad \dots (31)$$

[0113] 式(31)の括弧内における第一項 $(w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)) / (t^{(n-1)}_j - t')$ は、第 $n-1$ 層の j 番目のニューロンの発火の寄与によるものである。第二項 $(\sum_{j' \neq j} w^{(n)}_{ij'} \theta(t - t^{(n-1)}_{j'}) / (t^{(n-1)}_j - t'))$ は、第 $n-1$ 層の j 番目以外のニューロンの発火の寄与によるものである。発火確率密度 $R_{\text{linear}}(t)$ の変化 $\delta R_{\text{linear}}(t)$ は、 $R_{\text{linear}}(t)$ の傾き α の変化 $\delta \alpha$ と考えることができる。

式(31)の括弧内は傾き α を示しており、前段層のニューロンの発火時間 $t^{(n-1)}_j$ が $\delta t^{(n-1)}_j$ だけ変化したときに変化する部分は第 $n-1$ 層の j 番目のニューロンの発火の寄与による第一項のみである。すなわち、式(32)に示すように傾きが変化する。

[0114] [数32]

$$\begin{aligned}
 \alpha + \delta \alpha &= \frac{\sum_{j' \neq j} w^{(n)}_{ij'} \theta(t - t^{(n-1)}_{j'})}{t^{(n-1)}_j - t'} + \frac{w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)}{t^{(n-1)}_j - t' + \delta t^{(n-1)}_j} \\
 &= \frac{(t^{(n-1)}_j - t' + \delta t^{(n-1)}_j) \sum_{j' \neq j} w^{(n)}_{ij'} \theta(t - t^{(n-1)}_{j'}) + (t^{(n-1)}_j - t') w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)}{(t^{(n-1)}_j - t') (t^{(n-1)}_j - t' + \delta t^{(n-1)}_j)} \\
 &= \frac{(t^{(n-1)}_j - t' + \delta t^{(n-1)}_j) \sum_{j'} w^{(n)}_{ij'} \theta(t - t^{(n-1)}_{j'}) - \delta t^{(n-1)}_j w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)}{(t^{(n-1)}_j - t') (t^{(n-1)}_j - t' + \delta t^{(n-1)}_j)} \\
 &= \alpha - \frac{\delta t^{(n-1)}_j w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)}{(t^{(n-1)}_j - t') (t^{(n-1)}_j - t' + \delta t^{(n-1)}_j)} \\
 &= \alpha - \frac{\delta t^{(n-1)}_j w^{(n)}_{ij} \theta(t - t^{(n-1)}_j)}{(t^{(n-1)}_j - t')^2} \\
 &\quad \dots (32)
 \end{aligned}$$

[0115] また、式(19)から式(33)が得られる。

[0116] [数33]

$$\frac{\delta t^{(n)}_i}{\delta \alpha} \approx \frac{\partial t^{(n)}_i}{\partial \alpha} \approx \frac{\partial}{\partial \alpha} \left(t' + \frac{1}{\alpha} \right) = -\frac{1}{\alpha^2} \quad \dots (33)$$

[0117] これにより、偏微分 $\partial t^{(n)}_i / \partial t^{(n-1)}_j$ は、式(34)のように近似で

きる。

[0118] [数34]

$$\begin{aligned} \frac{\partial t_i^{(n)}}{\partial t_j^{(n-1)}} &\approx \frac{\delta t_i^{(n)}}{\delta t_j^{(n-1)}} = \frac{\delta \alpha}{\delta t_j^{(n-1)}} \frac{\delta t_i^{(n)}}{\delta \alpha} \\ &\approx -\frac{w_{ij}^{(n)} \theta(t_i^{(n)} - t_j^{(n-1)})}{(t_j^{(n-1)} - t')^2} \cdot -\frac{1}{\alpha^2} = \frac{w_{ij}^{(n)} \theta(t_i^{(n)} - t_j^{(n-1)})}{\alpha^2 (t_j^{(n-1)} - t')^2} \\ &\quad \dots (34) \end{aligned}$$

[0119] ここで定数 τ を式 (35) のようにする。

[0120] [数35]

$$\tau = (t_j^{(n-1)} - t') \quad \dots (35)$$

[0121] τ を用いて、式 (36) が得られる。

[0122] [数36]

$$\frac{\partial t_i^{(n)}}{\partial t_j^{(n-1)}} \approx \frac{w_{ij}^{(n)}}{\alpha^2 \tau^2} \theta(t_i^{(n)} - t_j^{(n-1)}) \quad \dots (36)$$

[0123] (学習則の具体例)

以上より、ニューラルネットワーク装置 100 における任意の層の重みの近似学習則が導出できるが、以下で、具体例として、第 N 層の学習則と、第 N-1 層の学習則について記す。

出力層の学習則は、式 (37) のようになる。

[0124] [数37]

$$\begin{aligned} \Delta w_{ij}^{(n)} &= -\eta_0^{(n)} \frac{\partial L_i}{\partial w_{ij}^{(n)}} \approx \frac{\eta_0^{(n)}}{\alpha} (t_i^{(n)} - t_i^{(t)}) \theta(t_i^{(n)} - t_j^{(n-1)}) \\ &= \eta^{(n)} (t_i^{(n)} - t_i^{(t)}) \theta(t_i^{(n)} - t_j^{(n-1)}) \\ &\quad \dots (37) \end{aligned}$$

[0125] ここで、 $\eta^{(n)}$ は、式（３８）のように示される。

[0126] [数38]

$$\eta^{(n)} = \frac{\eta_0^{(n)}}{\alpha} \quad \dots (38)$$

[0127] $\eta^{(n)}$ は、学習率を示す。ここで、式（３８）に示すように学習率 $\eta^{(n)}$ と発火確率密度の傾き α との組み合わせを用いて、学習率 $\eta^{(n)}$ を再定義する。式（３８）では、発火確率密度の傾き α を定数として扱う。

学習処理部３００は、式（３７）に基づいて出力層のニューロンモデル部１２１への入力に対する重み $w^{(N)}_{ij}$ を更新することで、出力層の学習を行わせる。上述したように、重み $w^{(N)}_{ij}$ は、第 $N-1$ 層の j 番目のニューロンモデル部１２１と、第 N 層の i 番目のニューロンモデル部１２１との結合の強さを示す。なお、出力層なので、式（３７）で $n=N$ と読み替える。

隠れ層の学習則（重みの更新則）の具体例は、式（３９）のようになる。

[0128] [数39]

$$\begin{aligned} \Delta w_j^{(n-1)} &= -\eta_0^{(n-1)} \frac{\partial L}{\partial w_{jk}^{(n-1)}} = -\eta_0^{(n-1)} \sum_i (t_i^{(n)} - t_i^{(t)}) \frac{\partial t_j^{(n-1)}}{\partial w_{jk}^{(n-2)}} \frac{\partial t_i^{(n)}}{\partial t_j^{(n-1)}} \\ &= -\eta_0^{(n-1)} \frac{\partial t_j^{(n-1)}}{\partial w_{jk}^{(n-2)}} \sum_i (t_i^{(n)} - t_i^{(t)}) \frac{\partial t_i^{(n)}}{\partial t_j^{(n-1)}} \\ &= \frac{\eta_0^{(n-1)}}{\alpha} \theta(t_j^{(n-1)} - t_k^{(n-2)}) \sum_i (t_i^{(n)} - t_i^{(t)}) \frac{w_{ij}^{(n)}}{\alpha^2 \tau^2} \theta(t_i^{(n)} - t_j^{(n-1)}) \\ &= \frac{\eta_0^{(n-1)}}{\alpha^3 \tau^2} \theta(t_j^{(n-1)} - t_k^{(n-2)}) \\ &= \eta^{(n-1)} \theta(t_j^{(n-1)} - t_k^{(n-2)}) \sum_i (t_i^{(n)} - t_i^{(t)}) w_{ij}^{(n)} \theta(t_i^{(n)} - t_j^{(n-1)}) \\ &\quad \dots (39) \end{aligned}$$

[0129] なお、 $\eta^{(n-1)}$ について、式（４０）とした。

[0130]

[数40]

$$\eta^{(n-1)} = \frac{\eta_0^{(n-1)}}{\alpha^3 \tau^2} \quad \dots (40)$$

[0131] 学習処理部300は、式(39)に基づいて、第n層（ここでは隠れ層とする）のニューロンモデル部121への入力に対する重み $w^{(n)}_{ij}$ を更新することで、隠れ層の学習を行わせる。上述したように、重み $w^{(n)}_{ij}$ は、第n-1層のj番目のニューロンモデル部121と、第n層のi番目のニューロンモデル部121との結合の強さを示す。

[0132] 図7は、ネットワークの重みの更新則の例を示す図である。図7は、実施形態に係る更新則を、第N層、第N-1層のそれぞれについて、Spike Propの場合の例と対比して表形式で示している。

図7の例では、出力層、隠れ層の何れに関しても、実施形態に係るアルゴリズムのほうが、Spike Propの場合よりも簡単な式で示されている。この点で、実施形態に係るアルゴリズムを用いることで、Spike Propの場合よりも、ネットワークの重みの更新処理（すなわち、ニューラルネットワークの学習処理）を比較的簡単なものとすることができる。

実施形態に係るアルゴリズムをソフトウェア的に実行する場合、ネットワークの重みの更新処理が比較的簡単なことで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済む。また、実施形態に係るアルゴリズムをハードウェア的に実行する場合、ネットワークの重みの更新処理が比較的簡単なことで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済むことに加えて、ハードウェアの回路面積が比較的小さくて済む。

[0133] (シミュレーション例)

実施形態に係るニューラルネットワーク装置100のシミュレーション例を示す。手書き文字データセットであるMNISTデータセットを対象に、実施形態に係るモデル（式(4)参照）を用いて、実施形態に係る学習アルゴリズム（図7参照）を用いて学習させた。MNISTデータセットは28

× 28画素の画像である784次元ベクトルと正解のスカラー値とが、訓練用として60000個ずつ与えられ、テスト用として同様の784次元ベクトルが10000個与えられている。

シミュレーションでは、ニューラルネットワークの重みを訓練用データを用いて更新し、テストデータを用いて性能を評価する。なお、テストデータを用いて重みは更新しない。

[0134] シミュレーションで用いたネットワークは三層であり、第一層は169個の入力用スパイキングニューロン、第二層と第三層はそれぞれ、500個、10個のスパイキングニューロン（式（4）参照）で構成されている。

なお、入力のスパイキングニューロンは、入力データの28×28画素の画像データを畳み込みによる前処理を行い、13×13の169画素に削減している。これにより、データ量を削減し、効率的なシミュレーションが可能になる。

また、一枚の画像毎に重みを更新する行うオンライン学習を行った。

また、図7に記載しているSpikePropアルゴリズムを用いたシミュレーションも行い、性能の比較を行った。

[0135] 図8は、シミュレーション結果を示す図である。線L31は、SpikePropアルゴリズムを用いた場合のシミュレーション結果を示している。線L32は上述した近似アルゴリズムを用いた場合の結果（ニューラルネットワーク装置100のシミュレーション結果）を示している。図8の横軸はエポック数を示し、縦軸はテスト時の分類エラー率を示す。

図8を参照すると、エポック数が大きくなるごとに、SpikePropアルゴリズム、近似アルゴリズムの両方において分類エラー率が減少していることがわかる。

[0136] 分類エラー率は、SpikePropアルゴリズムにおいて、3.8%であり、近似アルゴリズムで4.9%であった。このように、近似アルゴリズムを用いても、分類エラー率はほとんど変わらないことがわかる。

[0137] （実施形態に係る出力層の学習（2））

ニューラルネットワークシステム 1 における他の学習アルゴリズムについて説明する。

損失関数の重みによる微分は、式 (4 1) のようになる。

[0138] [数41]

$$\frac{\partial L_i}{\partial w_{ij}^{(l)}} = \frac{\partial L}{\partial t_i^{(l)}} \cdot \frac{\partial t_i^{(l)}}{\partial w_{ij}^{(l)}}, \quad \frac{\partial L}{\partial t_i^{(l)}} = \sum_s \frac{\partial L}{\partial t_s^{(l+1)}} \frac{\partial t_s^{(l+1)}}{\partial t_i^{(l)}} \quad \dots (4 1)$$

[0139] 式 (4 1) における右辺の二つの項 ($\partial t_i^{(l)} / \partial w_{ij}^{(l)}$ および $\partial t_s^{(l+1)} / \partial t_i^{(l)}$) を、膜電位の時間発展方程式を用いて線形近似し、簡易な学習則を導く。

上述したように、 $w_{ij}^{(l)}$ は、 $l-1$ 番目の層の j 番目のニューロンから、 l 番目の層の i 番目のニューロンへとつながる結合の強さ (重み) を示す。

$t_i^{(l)}$ は、 l 番目の層の i 番目のニューロンの発火タイミングを示す。

式 (4 1) に示される偏微分「 $\partial t_i^{(l)} / \partial w_{ij}^{(l)}$ 」と「 $\partial t_s^{(l+1)} / \partial t_i^{(l)}$ 」とを求めることにより、学習則の導出が可能である。これらは Spike Prop 法により、式 (4 2) のように計算できる。

[0140] [数42]

$$\frac{\partial t_i^{(l)}}{\partial w_{ij}^{(l)}} = -\frac{t_i^{(l)} - t_j^{(l-1)}}{\sum_s w_{is}^{(l)}}, \quad \frac{\partial t_k^{(l+1)}}{\partial t_j^{(l)}} = \frac{w_{Kj}^{(l+1)}}{\sum_s w_{ks}^{(l+1)}} \quad \dots (4 2)$$

[0141] 但し、式 (4 2) に示される 2 式のいずれも、右辺の分母の和 ($\sum_s$) では、対象とする重みと結合している前層のニューロンが後層のニューロンよりも早く発火した場合のみ和をとっている。この分母を平均場的近似することで、学習時に必要なパラメータの個数を大きく減ずることが可能となる。

[0142] まず、 $\partial t_i^{(l)} / \partial w_{ij}^{(l)}$ について説明する。

図 9 は、重みの変化に伴って膜電位が変化する様子を表す図である。図 9 は、重み $w_{ij}^{(l)}$ が、 $w_{ij}^{(l)} + \Delta W$ へと変化すると、時刻 $t_i^{(l)}$ における膜電位 $v_i^{(l)}$ が、 V_{th} から $V_{th} + \Delta V$ へと変化する様子を表している。

図9の横軸は時刻を示し、縦軸は膜電位を示す。線L41は、重み $w^{(l)}_{ij}$ が変化しないときの膜電位の時間発展の例を示す。線L42は、重み $w^{(l)}_{ij}$ が変化したときの膜電位の時間発展の例を示す。線L43は、重み $w^{(l)}_{ij}$ が変化したときの膜電位の時間発展の直線近似を示す。線L43によれば、発火時刻の近似解は、図9に示す時刻 $t_i^{(l)}$ となる。

前述の膜電位の変位 ΔV は、図9に示すように、式(43)のように導ける。

[0143] [数43]

$$\Delta V = \Delta W(t_i^{(l)} - t_j^{(l-1)}) \quad \dots (43)$$

[0144] すると、l層のj番目のニューロンに初めて発火が伝わった時刻 $\tau_j^{(l)}$ と発火の閾値 V_{th} とを用いることで、膜電位 $v^{*(l)}_i(t)$ の時間発展を、時間に対して線形近似することが可能である。この近似の結果、膜電位の時間発展の式は、式(44)のように導出できる。

[0145] [数44]

$$v_i^{*(l)}(t) = \frac{V_{th} + \Delta V}{t_i^{(l)} - \tau_i^{(l)}}(t - \tau_i^{(l)}) \quad \dots (44)$$

[0146] この近似下での発火タイミング $t_i^{*(l)}$ は、式(45)を解くことにより導出できる。

[0147] [数45]

$$v_i^{*(l)}(t_i^{*(l)}) = V_{th} \quad \dots (45)$$

[0148] 導出される式は、式(46)のようになる。

[0149] [数46]

$$t_i^{*(l)} = \tau_i^{(l)} + \frac{V_{th}}{V_{th} + \Delta V}(t_i^{(l)} - \tau_i^{(l)}) \quad \dots (46)$$

[0150] これにより、 $\partial t^{(l)}_i / \partial w^{(l)}_{ij}$ について、 $(t_i^{*(l)} - t_i^{(l)}) / \Delta W$ にて $\Delta W \rightarrow 0$ の極限を取ることにより、式(47)のように

偏微分の近似式を導出できる。

[0151] [数47]

$$\begin{aligned}
 \frac{t_i^{*(l)} - t_i^{(l)}}{\Delta W} &= \frac{1}{\Delta W} \left(\frac{V_{th} - (V_{th} + \Delta V)}{V_{th} + \Delta V} t_i^{(l)} + \frac{V_{th} + \Delta V - V_{th}}{V_{th} + \Delta V} \tau_i^{(l)} \right) \\
 &= \frac{1}{\Delta W} \left(\frac{-\Delta V}{V_{th} + \Delta V} t_i^{(l)} + \frac{\Delta V}{V_{th} + \Delta V} \tau_i^{(l)} \right) \\
 &= \frac{\Delta W (t_i^{(l)} - t_j^{(l-1)})}{\Delta W} \left(\frac{-1}{V_{th} + \Delta V} t_i^{(l)} + \frac{1}{V_{th} + \Delta V} \tau_i^{(l)} \right) \\
 &= (t_i^{(l)} - t_j^{(l-1)}) \left(\frac{-1}{V_{th} + \Delta V} t_i^{(l)} + \frac{1}{V_{th} + \Delta V} \tau_i^{(l)} \right) \\
 &\rightarrow - (t_i^{(l)} - t_j^{(l-1)}) \left(\frac{t_i^{(l)} - \tau_i^{(l)}}{V_{th}} \right) \\
 &\quad \dots (47)
 \end{aligned}$$

[0152] 次に、 $\partial t^{(l+1)}_j / \partial t^{(l)}_k$ の近似式を導出する。

図10は、発火タイミングの変化に伴って膜電位が変化する様子を表す図である。図10は、発火タイミングが $t^{(l)}_k$ から $t^{(l)}_k + \Delta T$ へと変化すると、時刻 $t^{(l+1)}_j$ における膜電位 $v^{(l+1)}_j$ は、 V_{th} から $V_{th} + \Delta V$ へと変化する様子を表している。

図10の横軸は時刻を示し、縦軸は膜電位を示す。線L51は、膜電位の時間発展の例を示す。線L52は、変化後の膜電位の時間発展を表している。線L53は、膜電位の時間発展の近似の例を示す。線L53によれば、発火時刻の近似解は、図10に示す時刻 $t_j^{*(l+1)}$ となる。

膜電位の変位 ΔV は、図10に示すように、 $-w^{(l+1)}_{jk} \Delta T$ と導ける。すると、先ほどと同様に $l+1$ 層の j 番目のニューロンに初めて発火が伝わった時刻 $t^{(l+1)}_j$ と発火閾値 V_{th} とを用いることで、膜電位 $v^{*(l+1)}_j(t)$ の時間発展を時間に対して線形に近似することが可能である。この平均近似の式は、式(48)のように導出できる。

[0153]

[数48]

$$v_j^{*(l+1)}(t) = \frac{V_{th} + \Delta V}{t_j^{(l+1)} - \tau_j^{(l+1)}} (t - \tau_j^{(l+1)}) \quad \dots (48)$$

[0154] この近似下での発火タイミング $t_j^{*(l+1)}$ は、式 (49) を解くことにより導出できる。

[0155] [数49]

$$v_j^{*(l+1)}(t_j^{*(l+1)}) = V_{th} \quad \dots (49)$$

[0156] 発火タイミング $t_j^{*(l+1)}$ は、式 (50) のように導出される。

[0157] [数50]

$$t_j^{*(l+1)} = \tau_j^{(l+1)} + \frac{V_{th}}{V_{th} + \Delta V} (t_j^{(l+1)} - \tau_j^{(l+1)}) \quad \dots (50)$$

[0158] これにより、 $\partial t_j^{(l+1)} / \partial t_k^{(l)}$ は、 $(t_j^{*(l+1)} - t_j^{(l+1)}) / \Delta T$ にて、 $\Delta T \rightarrow 0$ の極限を取ることで近似することが可能となり、式 (51) のように偏微分の近似式を導出できる。

[0159] [数51]

$$\begin{aligned} \frac{t_j^{*(l+1)} - t_j^{(l+1)}}{\Delta T} &= \frac{1}{\Delta T} \left(\frac{V_{th} - (V_{th} + \Delta V)}{V_{th} + \Delta V} t_j^{(l+1)} + \frac{V_{th} + \Delta V - V_{th}}{V_{th} + \Delta V} \tau_j^{(l+1)} \right) \\ &= \frac{-w_{jk}^{(l+1)} \Delta T}{\Delta T} \left(\frac{-1}{V_{th} + \Delta V} t_j^{(l+1)} + \frac{1}{V_{th} + \Delta V} \tau_j^{(l+1)} \right) \\ &= w_{jk}^{(l+1)} \left(\frac{t_j^{(l+1)} - \tau_j^{(l+1)}}{V_{th} + \Delta V} \right) \\ &\rightarrow w_{jk}^{(l+1)} \left(\frac{t_j^{(l+1)} - \tau_j^{(l+1)}}{V_{th}} \right) \quad \dots (51) \end{aligned}$$

[0160] 従って、 $\partial t_i^{(l)} / \partial w_{ij}^{(l)}$ は、式 (52) のように近似される。

[0161]

[数52]

$$\frac{\partial t_i^{(l)}}{\partial w_{ij}^{(l)}} \approx - \left(t_i^{(l)} - t_j^{(l-1)} \right) \left(\frac{t_i^{(l)} - \tau_j^{(l)}}{V_{th}} \right) \quad \dots (52)$$

[0162] また、 $\partial t^{(l+1)}_j / \partial t^{(l)}_k$ は、式(53)のように近似される。

[0163] [数53]

$$\frac{\partial t_j^{(l+1)}}{\partial t_k^{(l)}} \approx w_{jk}^{(l+1)} \left(\frac{t_j^{(l+1)} - \tau_j^{(l+1)}}{V_{th}} \right) \quad \dots (53)$$

[0164] 導出した $\partial t^{(l)}_i / \partial w^{(l)}_{ij}$ の近似式(式(52))および $\partial t^{(l+1)}_j / \partial t^{(l)}_k$ の近似式(式(53))を使うことで、他のニューロンモデルの情報の参照を大きく減じた学習則を導出することが可能である。

学習処理部300は、例えば、上述した式(41)に基づく学習の際に、式(52)および式(53)に示される近似を適用する。式(41)に基づく学習は、出力層の学習、隠れ層の学習の何れにも適用可能である。学習処理部300が、式(52)および式(53)に示される近似を式(41)に適用した学習にて、出力層および隠れ層のうち何れか一方の学習を行わせるようにしてもよいし、両方の学習を行わせるようにしてもよい。

[0165] 以上のように、ニューロンモデル部121は、漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンとして構成され、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでニューロンモデル部121自らの出力を示す。伝達処理部122は、ニューロンモデル部121の間の情報伝達を行う。

[0166] ここでいうニューラルネットワークの1回の処理は、ニューラルネットワークが一群の入力データに対して出力データを出力する処理である。例えば、ニューラルネットワークがパターンマッチングを行う場合、1回のマッチング処理がニューラルネットワークの1回の処理の例に該当する。

ニューラルネットワーク装置100によれば、ニューロンモデル部121

のリークをなくし、さらにすべてのニューロンモデル部 121 が一回以下しか発火しない条件の下で、ニューロンモデル部 121 を、ステップ関数を用いた比較的簡単なモデルとすることができる。

[0167] ニューロンモデル部 121 の処理をソフトウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済む。また、ニューロンモデル部 121 の処理をハードウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済むことに加えて、ハードウェアの回路面積が比較的小さくて済む。

また、ニューロンモデル部 121 によれば、リークを含まないモデルである点で、ニューロンが時定数を有しないモデルとなり、入力データの時定数に依存しなくなるため、認識精度が高い。

また、ニューロンモデル部 121 は、時間方式による点で、頻度方式による場合よりも消費電力が少なくて済む。

[0168] また、学習処理部 300 は、膜電位の時間発展の線形近似を用いて得られる、発火時刻の重みによる微分の近似、および、発火時刻の発火時刻による微分の近似のうち少なくとも何れか一方を適用した学習則を用いて、ニューラルネットワーク装置 100 の出力層および隠れ層のうち少なくとも何れか一方の学習を行わせる。

これにより、ニューラルネットワークシステム 1 では、出力層および隠れ層のうち少なくとも何れか一方の学習を、近似を用いた比較的簡単な処理にて実行することができる。

学習処理部 300 による学習のアルゴリズムをソフトウェア的に実行する場合、学習処理が比較的簡単なことで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済む。また、学習処理部 300 による学習のアルゴリズムをハードウェア的に実行する場合、学習処理が比較的簡単なことで、処理負荷が比較的軽くて済み、処理時間が比較

的短くて済み、消費電力が比較的小さくて済むことに加えて、ハードウェアの回路面積が比較的小さくて済む。

なお、発火時刻の重みによる微分は、発火時刻を重みで微分したものを意味する。発火時刻の発火時刻による微分は、あるニューロンモデル部 1 2 1 の発火時刻を、他のニューロンモデル部 1 2 1 の発火時刻で微分したものを意味する。

[0169] また、学習処理部 3 0 0 は、発火確率密度の傾きを用いて示される学習則を用いて、前記ニューラルネットワーク装置の出力層の学習を行わせる。

これにより、ニューラルネットワークシステム 1 では、発火確率密度の変化に基づいて発火時刻の変化を求めることができ、この点で、発火時刻の変化を比較的容易に求めることができる。

[0170] 次に、図 1 1 を参照して、本発明の実施形態の構成について説明する。

図 1 1 は、実施形態に係るニューラルネットワーク装置の構成例を示す図である。図 1 1 に示すニューラルネットワーク装置 1 0 は、ニューロンモデル部 1 1 と、伝達処理部 1 2 とを備える。

かかる構成にて、ニューロンモデル部 1 1 は、漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンとして構成され、ニューラルネットワークの 1 回の処理において高々 1 回発火して発火タイミングでニューロンモデル部 1 1 自らの出力を示す。伝達処理部 1 2 は、ニューロンモデル部 1 1 の間の情報伝達を行う。

[0171] ニューラルネットワーク装置 1 0 によれば、ニューロンモデル部 1 1 のリークをなくし、さらにすべてのニューロンモデル部 1 1 が一回以下しか発火しない条件の下で、ニューロンモデル部 1 1 を、ステップ関数を用いた比較的簡単なモデルとすることができる。

[0172] ニューロンモデル部 1 1 の処理をソフトウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済む。また、

ニューロンモデル部 11 の処理をハードウェア的に実行する場合、ニューロンモデルが比較的簡単なモデルとなることで、処理負荷が比較的軽くて済み、処理時間が比較的短くて済み、消費電力が比較的小さくて済むことに加えて、ハードウェアの回路面積が比較的小さくて済む。

また、ニューロンモデル部 11 によれば、リークを含まないモデルである点で、ニューロンが時定数を有しないモデルとなり、入力データの時定数に依存しなくなるため、認識精度が高い。

また、ニューロンモデル部 11 は、時間方式による点で、頻度方式による場合よりも消費電力が少なくて済む。

[0173] ニューラルネットワークシステム 1 の全部または一部、あるいは、ニューラルネットワーク装置 10 の全部または一部が、専用ハードウェアに実装されていてもよい。

図 12 は、少なくとも 1 つの実施形態に係る専用ハードウェアの構成例を示す概略ブロック図である。図 12 に示す構成で、専用ハードウェア 500 は、CPU 510 と、主記憶装置 520 と、補助記憶装置 530 と、インタフェース 540 とを備える。

[0174] 上述のニューラルネットワークシステム 1 が専用ハードウェア 500 に実装される場合、上述した各処理部（ニューラルネットワーク装置 100、ニューロンモデル部 121、伝達処理部 122、予測誤差算出部 200、学習処理部 300）の動作は、プログラム、もしくは回路の形式で専用ハードウェア 500 に記憶されている。

[0175] ニューラルネットワークシステム 1 の全部または一部、あるいは、ニューラルネットワーク装置 10 の全部または一部が、ASIC (application specific integrated circuit) に実装されていてもよい。

図 13 は、少なくとも 1 つの実施形態に係るコンピュータの構成例を示す概略ブロック図である。図 13 に示す構成で、ASIC 600 は、演算部 610 と、記憶装置 620 と、インタフェース 630 とを備える。また、演算部 610 と記憶装置 620 とは統一されていても（すなわち、一体的に構成

されていても) よい。

[0176] ニューラルネットワークシステム 1 の全部または一部、あるいは、ニューラルネットワーク装置 10 の全部または一部が、実装された A S I C は、C M O S などの電子回路により、その演算を実行する。各々の電子回路が、それぞれ独立に層内のニューロンを実装してもよいし、層内の複数のニューロンを実装してもよい。また、同様に、ニューロンを演算する回路が、それぞれ、ある層の演算のみに用いられてもよいし、複数の層の演算に用いられてもよい。

[0177] ニューラルネットワーク装置 10 の全部または一部が A S I C に実装される場合、その A S I C は、特定のものに限定されない。例えば、ニューラルネットワーク装置 10 の全部または一部が、C P U を有していない A S I C に実装されていてもよい。また、ニューラルネットワーク装置 10 の実装に用いられる記憶装置が、チップ上に分散して配置されていてもよい。

[0178] なお、ニューラルネットワークシステム 1 の機能の全部または一部を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより各部の処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、O S (Operating System) や周辺機器等のハードウェアを含むものとする。

また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、R O M (Read Only Memory) 、C D - R O M (Compact Disc Read Only Memory) 等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであっても良い。

[0179] 以上、この発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱し

ない範囲の設計等も含まれる。

[0180] この出願は、2019年3月20日に出願された日本国特願2019-052880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。

産業上の利用可能性

[0181] 本発明は、ニューラルネットワーク装置、ニューラルネットワークシステム、処理方法および記録媒体に適用してもよい。

符号の説明

[0182] 1 ニューラルネットワークシステム
10、100 ニューラルネットワーク装置
11、121 ニューロンモデル部（ニューロンモデル手段）
12、122 伝達処理部（伝達処理手段）
111 第1層
112 第2層
113 第3層
114 第4層
200 予測誤差算出部（予測誤差算出手段）
300 学習処理部（学習処理手段）

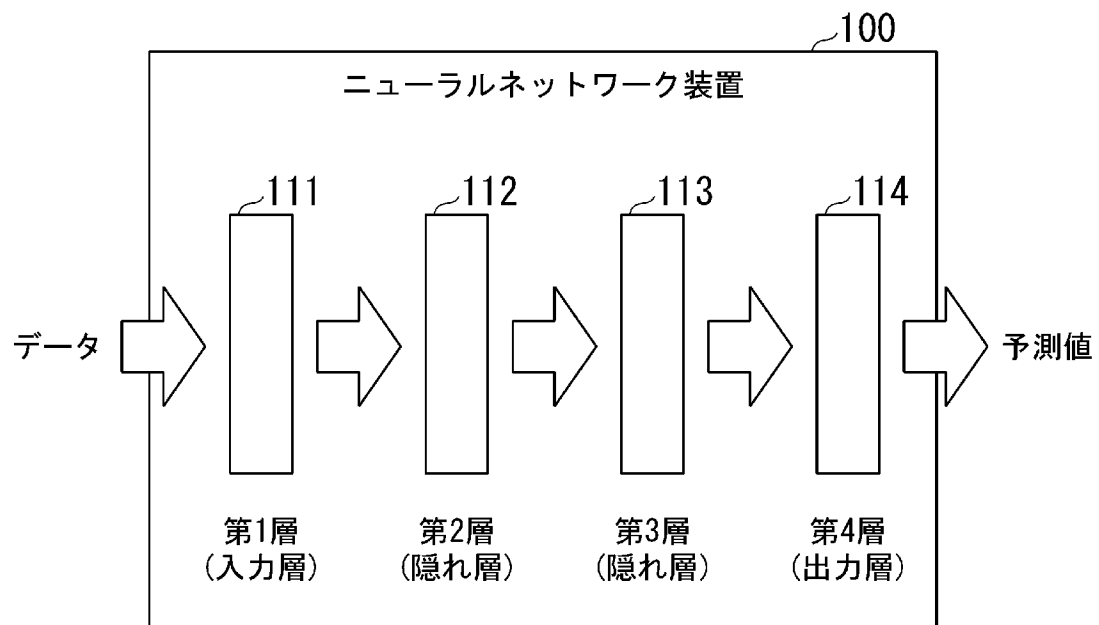
請求の範囲

- [請求項1] 漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンとして構成され、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでニューロンモデル手段自らの出力を示すニューロンモデル手段と、
前記ニューロンモデル手段の間の情報伝達を行う伝達処理手段と、
を備えるニューラルネットワーク装置。
- [請求項2] 請求項1に記載のニューラルネットワーク装置と、
膜電位の時間発展の線形近似を用いて得られる、発火時刻の重みによる微分の近似、および、発火時刻の発火時刻による微分の近似のうち少なくとも何れか一方を適用した学習則を用いて、前記ニューラルネットワーク装置の出力層および隠れ層のうち少なくとも何れか一方の学習を行わせる学習処理手段と、
を備えるニューラルネットワークシステム。
- [請求項3] 前記学習処理手段は、発火確率密度の傾きを用いて示される学習則を用いて、前記ニューラルネットワーク装置の出力層の学習を行わせる、
請求項2に記載のニューラルネットワークシステム。
- [請求項4] 漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後電流がステップ関数を用いて表されるスパイキングニューロンであって、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでスパイキングニューロン自らの出力を示すスパイキングニューロンの動作を実行する工程と、
前記スパイキングニューロン間の情報伝達を行う工程と、
を含む処理方法。
- [請求項5] A S I Cに、
漏れなし積分発火型のスパイキングニューロン、かつ、シナプス後

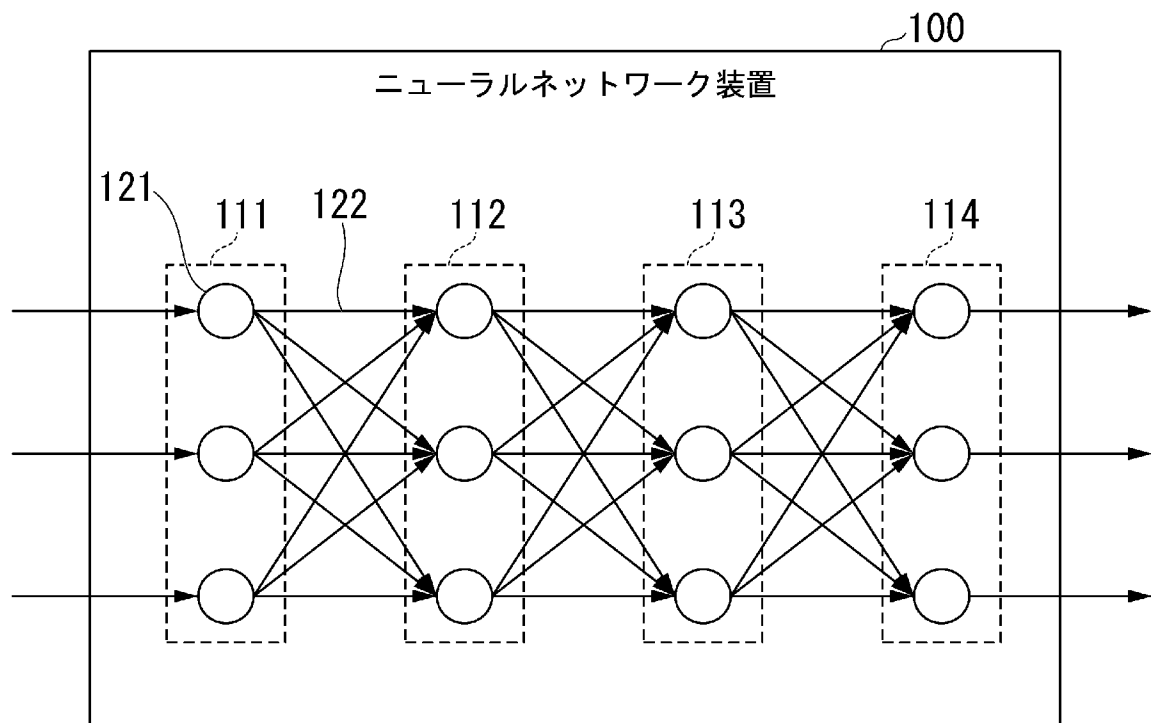
電流がステップ関数を用いて表されるスパイキングニューロンであって、ニューラルネットワークの1回の処理において高々1回発火して発火タイミングでスパイキングニューロン自らの出力を示すスパイキングニューロンの動作を実行する工程と、

前記スパイキングニューロン間の情報伝達を行う工程と、
を実行させるためのプログラムを記憶する記録媒体。

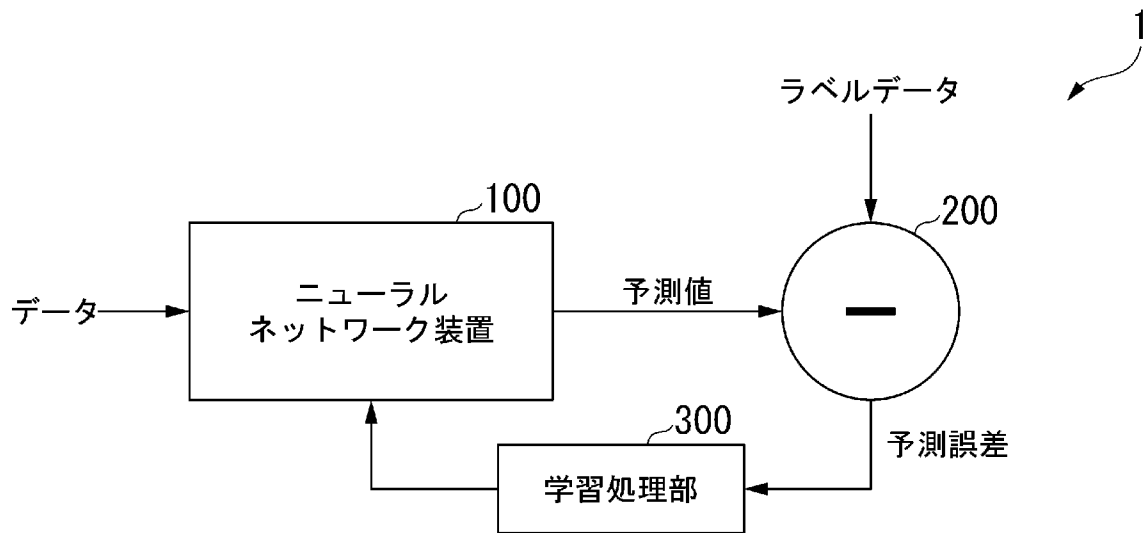
[図1]



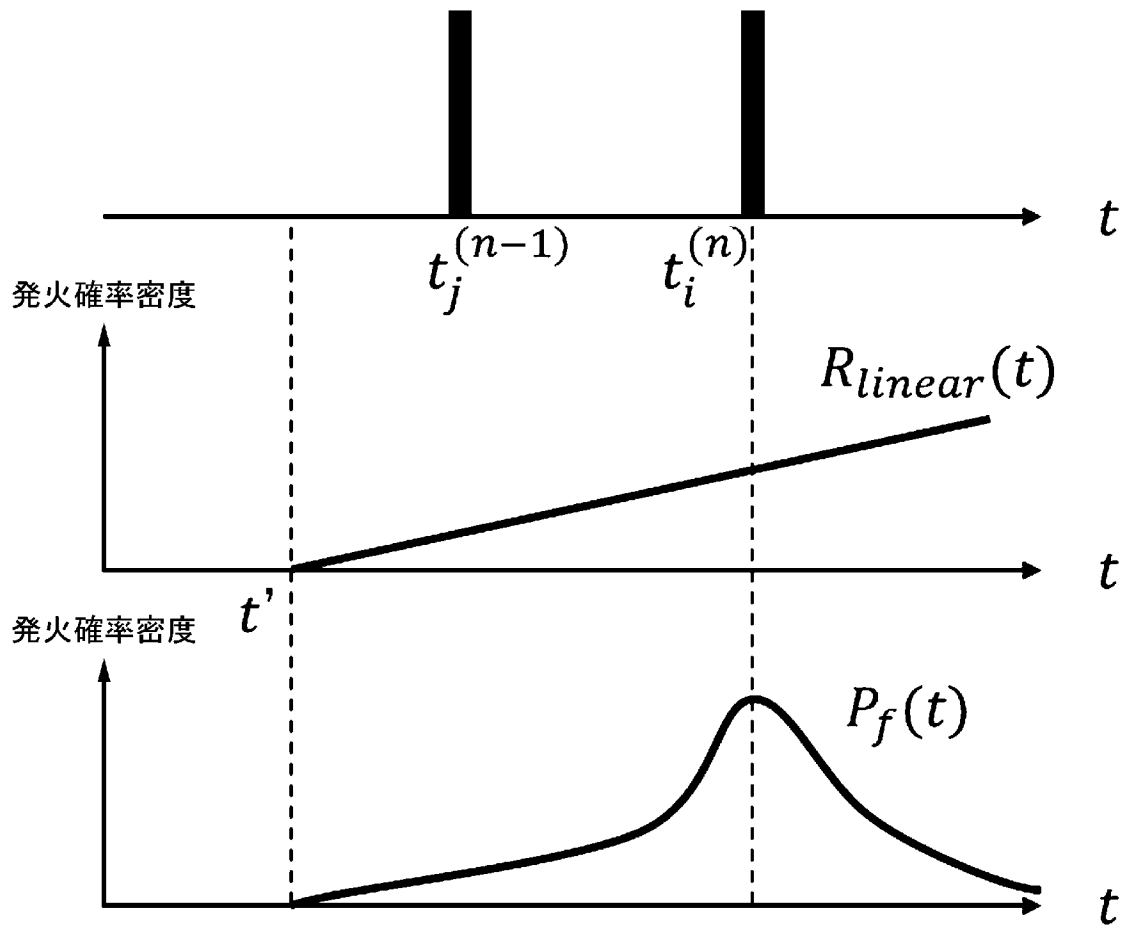
[図2]



[図3]

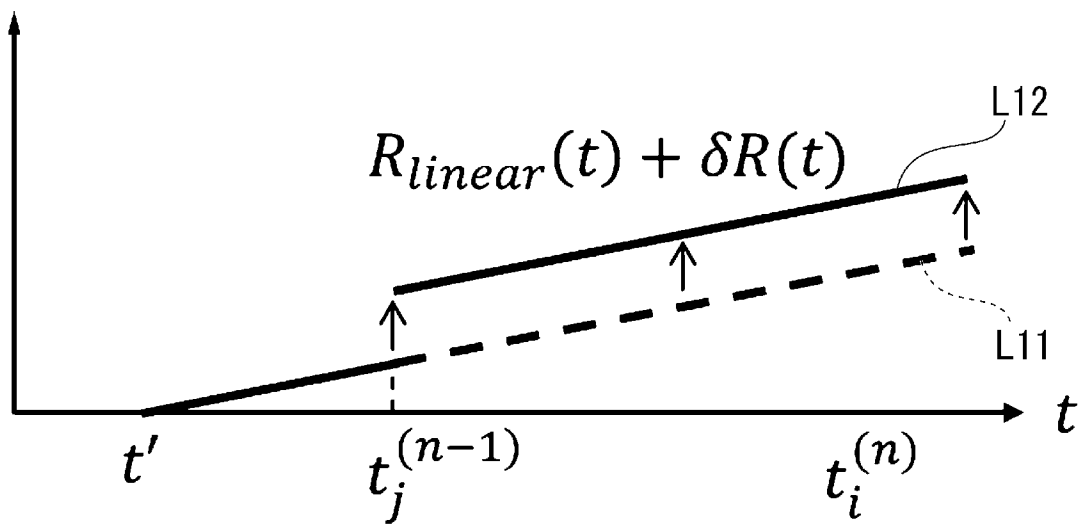


[図4]



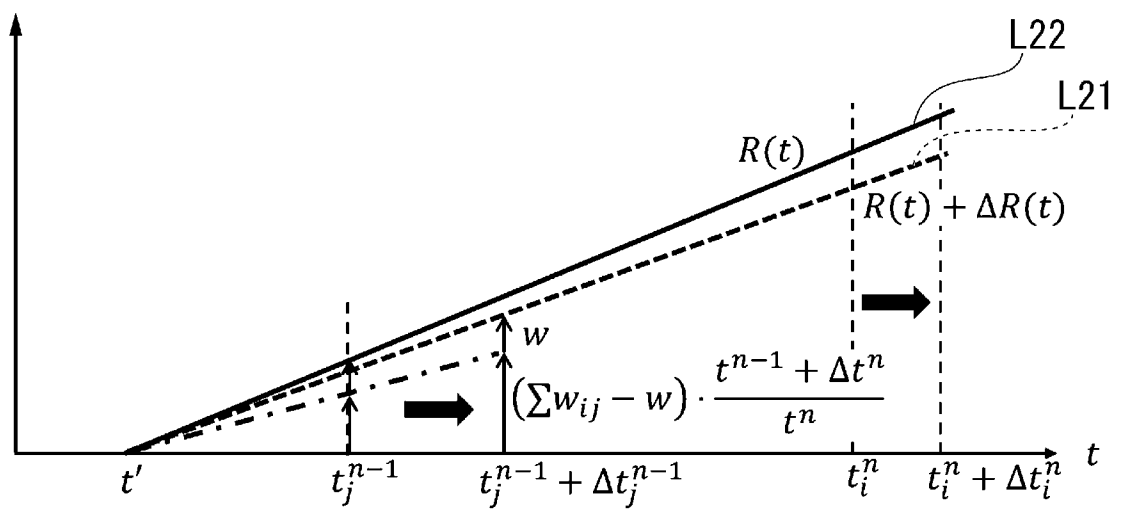
[図5]

発火確率密度



[図6]

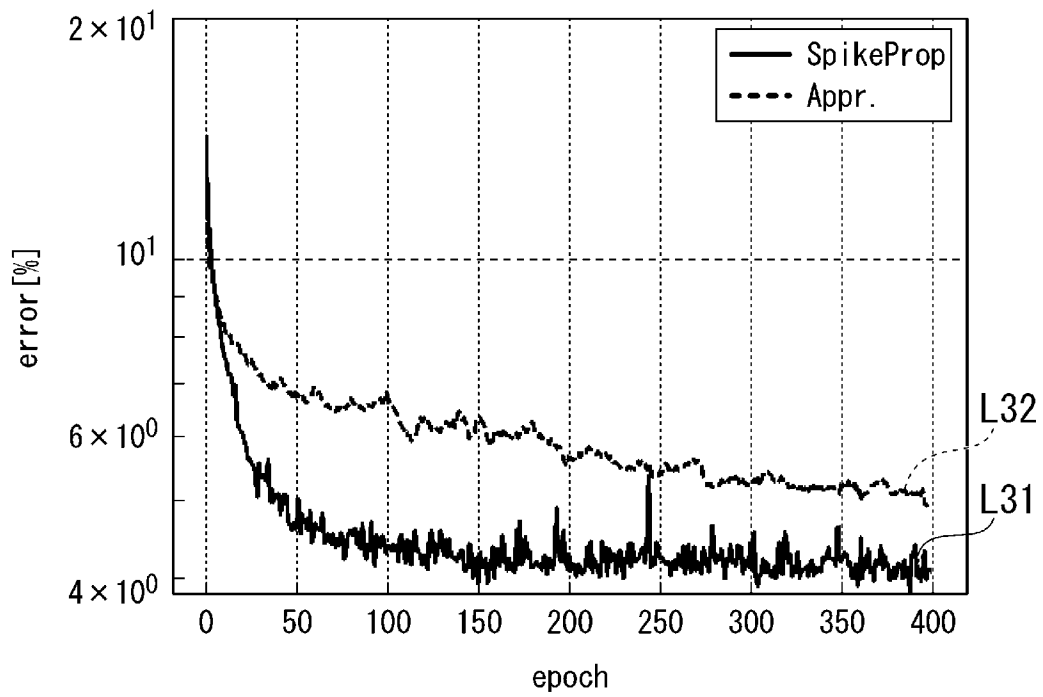
発火確率密度



[図7]

	SpikeProp	実施形態に係るアルゴリズム
出力層 (Δw_{ij}^n)	$\frac{\eta\left(t_i^{(n)}-t_i^{(t)}\right)\left(t_i^{(n)}-t_j^{(n)}\right)\theta\left(t_i^{(n)}-t_j^{(n-1)}\right)}{\sum_{j'} w_{ij'}^{(n)} \theta\left(t_i^{(n)}-t_{j'}^{(n-1)}\right)}$	$\eta^{(n)}\left(t_i^{(n)}-t_i^{(t)}\right)\theta\left(t_i^{(n)}-t_j^{(n-1)}\right)$
隠れ層 (Δw_{jk}^{n-1})	$\frac{\eta\left(t_j^{(n-1)}-t_k^{(n-2)}\right)\theta\left(t_j^{(n-1)}-t_k^{(n-2)}\right)}{\sum_{k'} w_{jk'}^{(n-1)} \theta\left(t_j^{(n-1)}-t_{k'}^{(n-2)}\right)}$ $\cdot \frac{\sum_i\left(t_i^{(n)}-t_i^{(t)}\right) w_{ij}^{(n)} \theta\left(t_i^{(n)}-t_j^{(n-1)}\right)}{\sum_{j'} w_{ij'}^{(n)} \theta\left(t_i^{(n)}-t_{j'}^{(n-1)}\right)}$	$\eta^{(n-1)}\theta\left(t_j^{(n-1)}-t_k^{(n-2)}\right)$ $\cdot \sum_i\left(t_i^{(n)}-t_i^{(t)}\right) w_{ij}^{(n)} \theta\left(t_i^{(n)}-t_j^{(n-1)}\right)$

[図8]



[図9]

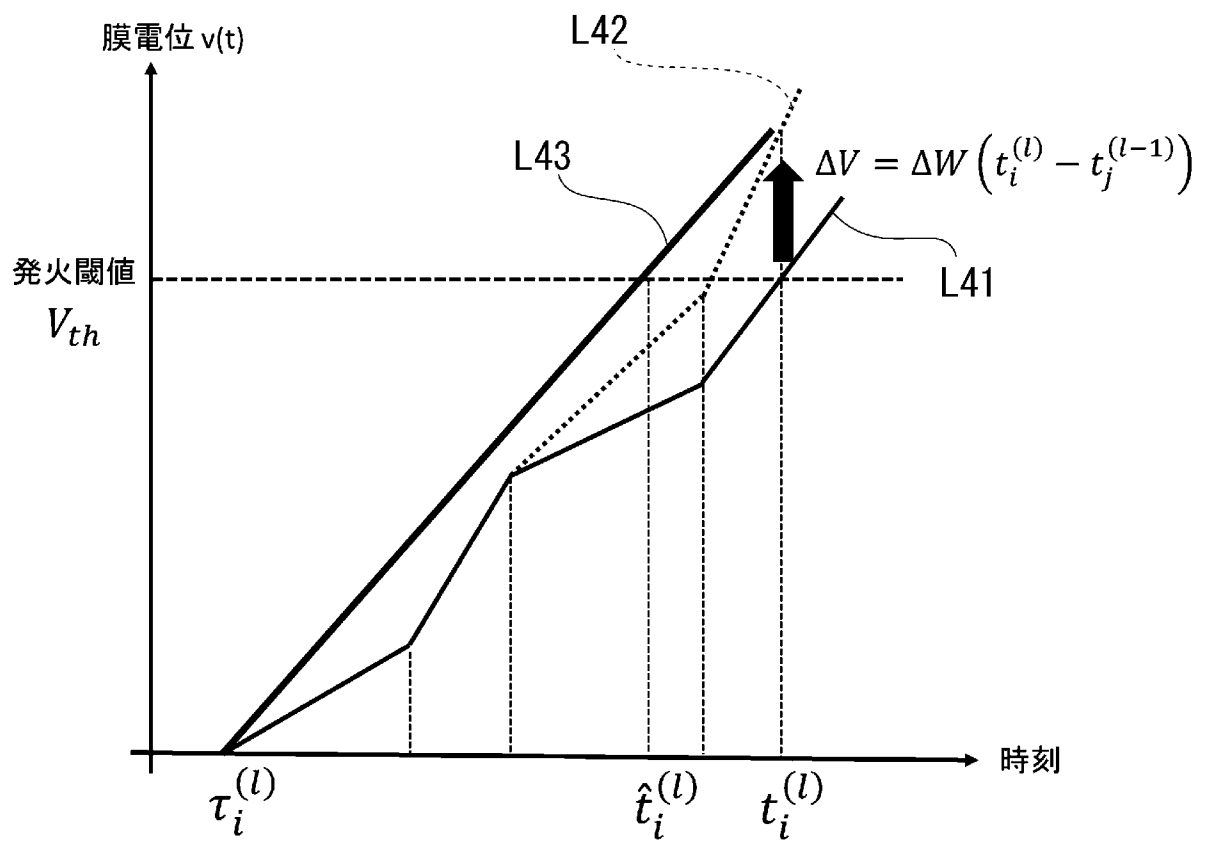


Diagram illustrating a neural network device (10). The device consists of two neuron model units (11) connected by a transmission processing unit (12). The overall structure is labeled as the neural network device (10).

```

graph LR
    subgraph 10 [ニューラルネットワーク装置]
        direction LR
        11L[ニューロンモデル部] --- 12[伝達処理部] --- 11R[ニューロンモデル部]
    end

```

```
graph TD; CPU[CPU 510] --- Bus; Bus --- MemMain[主記憶装置 520]; Bus --- MemAux[補助記憶装置 530]; Bus --- Intf[インタフェース 540];
```

500

510

CPU

520

主記憶装置

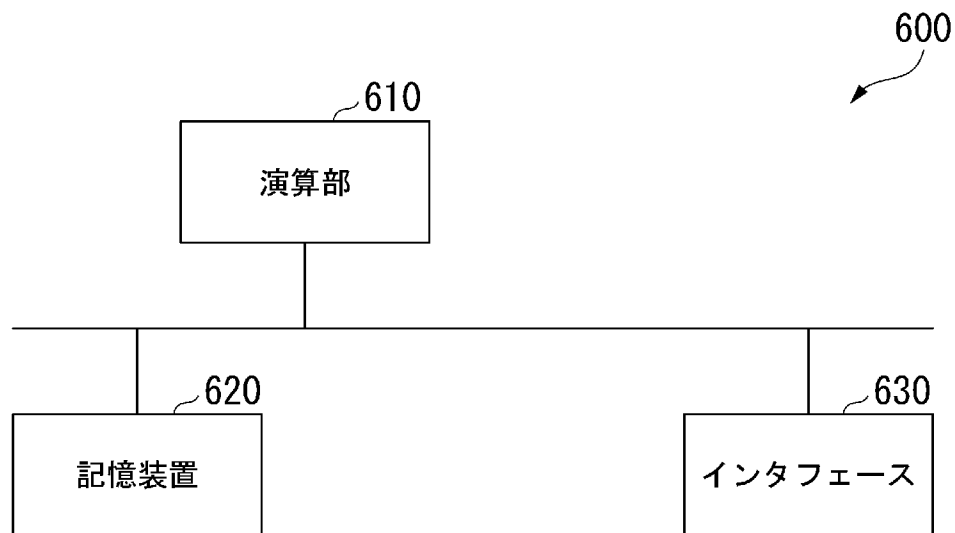
530

補助記憶装置

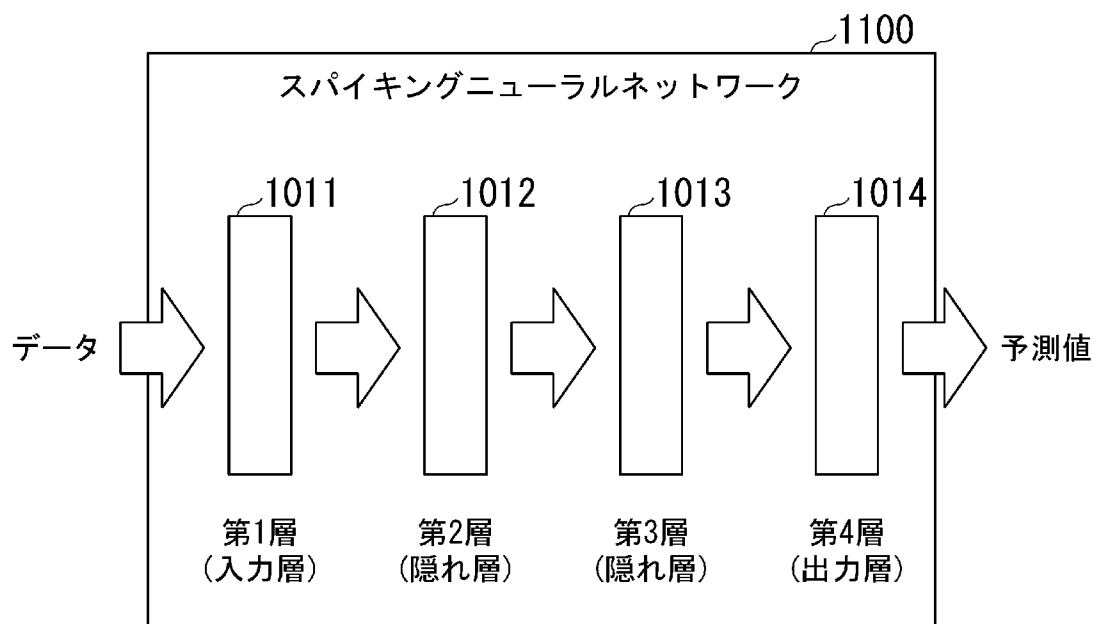
540

インタフェース

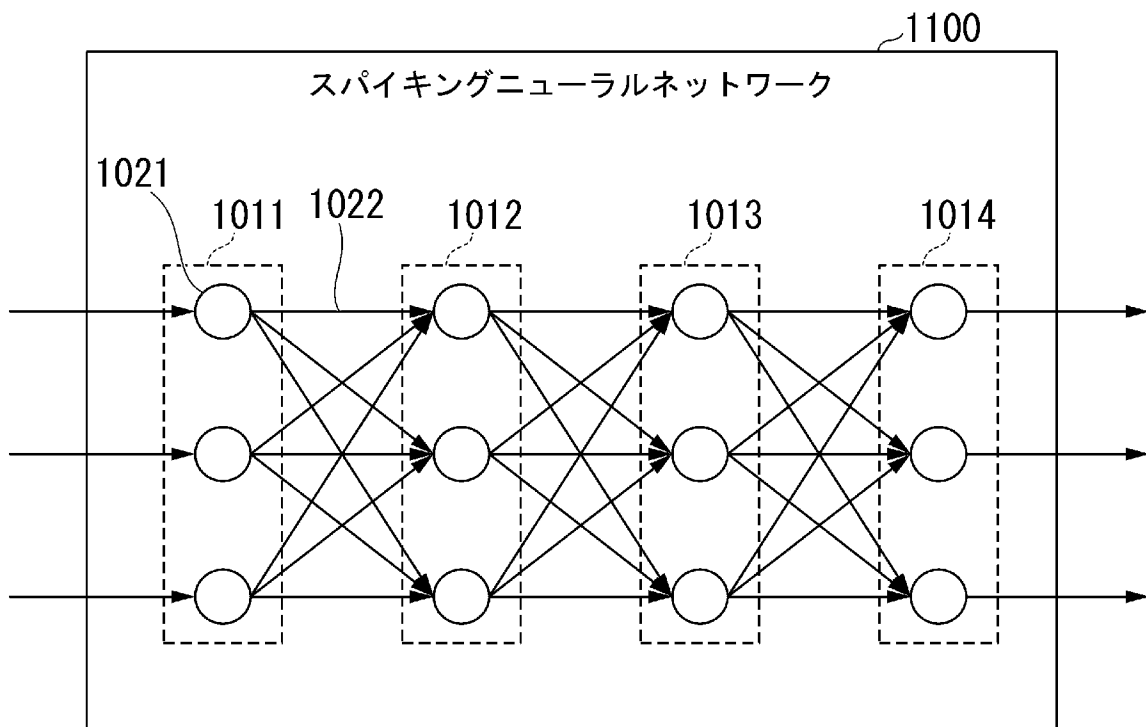
[図13]



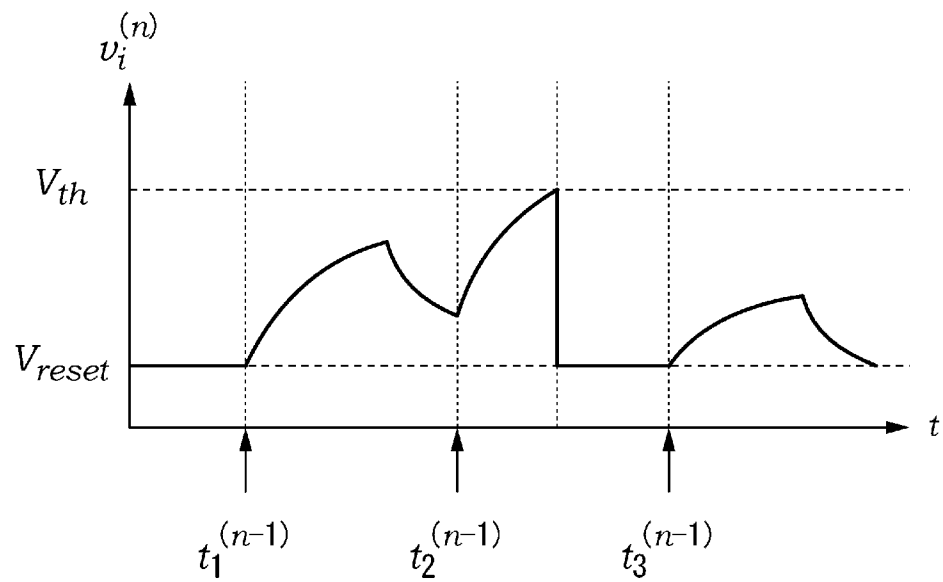
[図14]



[図15]



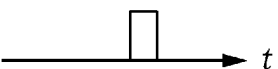
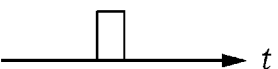
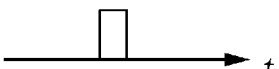
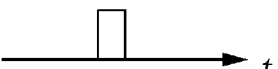
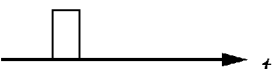
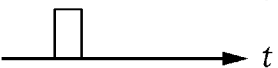
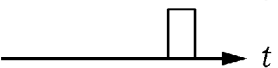
[図16]



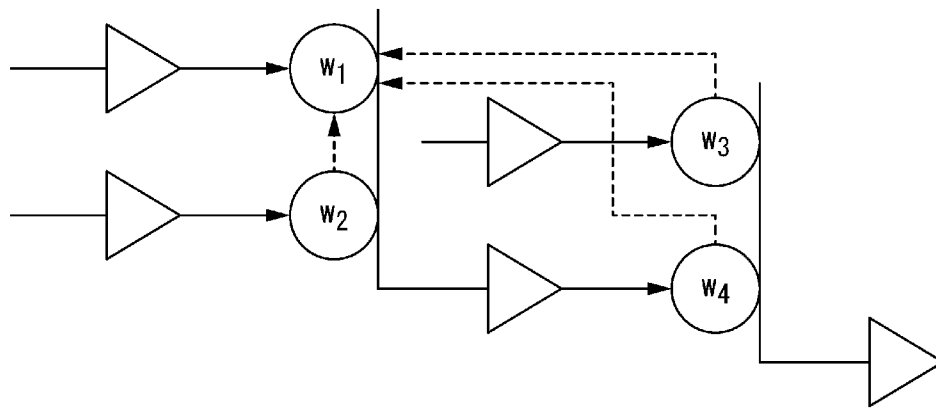
[図17]

情報	頻度方式	時間方式
“1”		
“3”		
“5”		

[図18]

出力ニューロン	ケース1	ケース2	ケース3
0番目			
1番目			
2番目			
予測値	0	2	1

[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/011909

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06N3/04 (2006.01) i

FI: G06N3/04190

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. G06N3/04

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 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

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	MOSTAFA, H., "Supervised learning based on temporal coding in spiking neural networks", arXiv, pp. 1-12 [online], 2017.08.16, [retrieved on 13 May 2020], Internet: <URL:https://arxiv.org/abs/1606.08165>, particularly, sections 2, 3	1, 4-5 2-3

☐ Further documents are listed in the continuation of Box C.	☐ See patent family annex.
---	---

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 13.05.2020	Date of mailing of the international search report 02.06.2020
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）） G06N 3/04(2006.01)i FI: G06N3/04 190														
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）） G06N3/04														
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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2020年													
日本国実用新案登録公報	1996 - 2020年													
日本国登録実用新案公報	1994 - 2020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	MOSTAFA, Hesham, "Supervised learning based on temporal coding in spiking neural networks", arXiv, pp.1-12 [online], 2017.8.16, [検索日 2020.5.13], インターネット: <URL: https://arxiv.org/abs/1606.08165 >, 特にSection2-3	1,4-5												
A	Section2-3	2-3												
☐ 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。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>"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抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>"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</td> <td>"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</td> </tr> <tr> <td>"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</td> <td>"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</td> </tr> <tr> <td>"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）</td> <td>"&" 同一パテントファミリー文献</td> </tr> <tr> <td>"O" 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>"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</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	"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抵触するものではなく、発明の原理又は理論の理解のために引用するもの	"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	"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	"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	"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	"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）	"&" 同一パテントファミリー文献	"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	
* 引用文献のカテゴリー	"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抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
"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	"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													
"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	"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													
"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）	"&" 同一パテントファミリー文献													
"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														
国際調査を完了した日 13.05.2020	国際調査報告の発送日 02.06.2020													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 杉浦 孝光 5B 4808 電話番号 03-3581-1101 内線 3545													