

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4931311号

(P4931311)

(45) 発行日 平成24年5月16日(2012.5.16)

(24) 登録日 平成24年2月24日(2012.2.24)

(51) Int.Cl.

F I

G 0 6 N 3 / 0 6 3 (2006.01)

G 0 6 N 3 / 0 6 3

請求項の数 18 (全 13 頁)

(21) 出願番号	特願2001-532486 (P2001-532486)	(73) 特許権者	593091474
(86) (22) 出願日	平成12年10月16日(2000.10.16)		ザ・ビクトリア・ユニバーシティ・オブ・
(65) 公表番号	特表2003-512683 (P2003-512683A)		マンチエスター
(43) 公表日	平成15年4月2日(2003.4.2)		イギリス国、エム・13・9・ビー・エル
(86) 国際出願番号	PCT/GB2000/003957		、マンチエスター、オックスフォード・ロ
(87) 国際公開番号	W02001/029766		ード(番地なし)
(87) 国際公開日	平成13年4月26日(2001.4.26)	(74) 代理人	100084146
審査請求日	平成19年9月26日(2007.9.26)		弁理士 山崎 宏
(31) 優先権主張番号	9924310.7	(74) 代理人	100100170
(32) 優先日	平成11年10月15日(1999.10.15)		弁理士 前田 厚司
(33) 優先権主張国	英国(GB)	(74) 代理人	100111039
			弁理士 前堀 義之
前置審査			
			最終頁に続く

(54) 【発明の名称】 ニューラルネットワーク素子

(57) 【特許請求の範囲】

【請求項1】

複数の入力部と、少なくとも一つのプロセッシング要素と、少なくとも一つ出力部と、それぞれ前記少なくとも一つのプロセッシング要素に対応するアドレスに値を記憶するデジタルメモリとを備える、ニューラルネットワーク素子であって、

前記少なくとも一つのプロセッシング要素は、他のニューラルネットワーク素子のプロセッシング要素からの入力信号にตอบสนองして前記デジタルメモリからの命令を含む値を受信するように構成され、かつ前記デジタルメモリから受信された値に応じて複数の操作の一つを実行するように命令され、

前記複数の操作の各々は、前記少なくとも一つのプロセッシング要素の活性レベルを変更し、前記活性レベルは前記プロセッシング要素の興奮度を示す内的値である、ニューラルネットワーク素子。

【請求項2】

前記操作は、前記少なくとも一つのプロセッシング要素の活性レベルへの前記値の加算である、請求項1に記載のニューラルネットワーク素子。

【請求項3】

前記操作は、少なくとも一つのプロセッシング要素の活性レベルの零への減衰であり、この減衰は前記活性レベルが負であれば前記活性レベルを増加させ、前記活性レベルが正であれば前記活性レベルを減少させることによるものである、請求項1に記載のニューラルネットワーク素子。

10

20

【請求項 4】

前記減衰の速度は前記デジタルメモリから受信された値に依存する、請求項 3 に記載のニューラルネットワーク素子。

【請求項 5】

前記減衰はクロックにより制御される、請求項 3 又は請求項 4 に記載のニューラルネットワーク素子。

【請求項 6】

前記操作は、前記少なくとも一つのプロセッシング要素による出力の生成と、その少なくとも一つのプロセッシング要素の活性レベルの予め定められた最小レベルへのリセットとを含む、請求項 1 に記載のニューラルネットワーク素子。

10

【請求項 7】

前記操作は、前記少なくとも一つのプロセッシング要素により出力を生成することなく、その少なくとも一つのプロセッシング要素の活性レベルをリセットすることを含む、請求項 1 に記載のニューラルネットワーク素子。

【請求項 8】

前記操作は、自動学習アルゴリズムに基づいて、前記少なくとも一つの活性レベルを前記デジタルメモリに保持された値に変更することを含む、請求項 1 に記載のニューラルネットワーク素子。

【請求項 9】

前記操作は、自動学習アルゴリズムに基づいて、前記デジタルメモリに保持された値を変更することを含む、請求項 1 又は請求項 8 に記載のニューラルネットワーク素子。

20

【請求項 10】

前記自動学習アルゴリズムは、ヘブ学習アルゴリズムを備える請求項 8 又は請求項 9 に記載のニューラルネットワーク素子。

【請求項 11】

前記ニューラルネットワーク素子により生成される出力は、前記出力を発生した前記プロセッシング要素の個々のアドレスの番号を含む、請求項 1 に記載のニューラルネットワーク素子。

【請求項 12】

前記ニューラルネットワーク素子は、前記プロセッシング要素から受信された信号パルスの時系列順序を決定し、各信号パルスを出力を発生したプロセッシング要素のアドレスを表す番号により表すように構成された調停及びエンコード手段を備える、請求項 11 に記載のニューラルネットワーク素子。

30

【請求項 13】

前記デジタルメモリに対する読出線アクセスを有する通常のプロセッサをさらに備える、請求項 1 に記載のニューラルネットワーク素子。

【請求項 14】

前記通常のプロセッサによる前記デジタルメモリへのアクセスは、前記ニューラルプロセッシング要素により要求されるアクセスに対して非同期で調停される、請求項 13 に記載のニューラルネットワーク素子。

40

【請求項 15】

多数のニューラルネットワーク素子が通常のプロセッサの並列ネットワークにより管理される、請求項 13 又は請求項 14 に記載のニューラルネットワーク素子。

【請求項 16】

前記デジタルメモリはランダムアクセスメモリである、請求項 1 に記載のニューラルネットワーク素子。

【請求項 17】

前記素子は非同期法で作動する請求項 1 に記載のニューラルネットワーク素子。

【請求項 18】

複数の入力部と、少なくとも一つのプロセッシング要素と、少なくとも一つの出力部と

50

、それぞれ前記少なくとも一つのプロセッシング要素に対応するアドレスに値を記憶するデジタルメモリとを備えるニューラルネットワーク素子进行操作する方法であって、

前記少なくとも一つのプロセッシング要素にそれぞれ対応する前記デジタルメモリのアドレスに値を記憶し、

他のニューラルネットワーク素子のプロセッシング要素からの入力信号に応答して前記少なくとも一つのプロセッシング要素に命令を含む値を伝送し、かつ

前記デジタルメモリから受信された値に応じて前記プロセッシング要素で複数の操作のうちの一つを実行し、

前記複数の操作の各々は、前記少なくとも一つのプロセッシング要素の活性レベルを変更し、前記活性レベルは前記プロセッシング要素の興奮度を示す内的値である、

10

ニューラルネットワーク素子进行操作する方法。

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、ニューラルネットワーク素子に関する。

【0002】

(背景技術)

ニューラルネットワークは、生物学的に着想された情報処理のための計算アーキテクチャである。通常のプログラム内蔵型コンピュータで実行される通常のアルゴリズムでは解くことが困難な問題を解くためにニューラルネットワークを使用することが多くなっている。一般的には、これらの問題は株式市場予測、画像認識、音声認識のようなパターンマッチング問題である。いくつかのニューラルネットワークの応用例は商業的に重要である。例えば、多くのラップトップパーソナルコンピュータのタッチパッドは、信頼性を高めるためにニューラルネットワークを使用している(ニューラルシステムは比較的ノイズ不感性である。)。

20

【0003】

ニューロンは、概して多入力、単一出力のデバイスである。1個のニューロンからの出力信号はそのニューロンに対する入力の重み付けされた総和の関数であり、以下の等式により表すことができる。

【0004】

30

【数1】

$$Y_i = f_i \left(\sum_j W_{ij} \cdot X_j - T_i \right)$$

【0005】

ここで、 X_j はニューロンへの入力(場合によっては他のニューロンからの)、 W_{ij} は入力に適用される荷重、以下の総和はニューロンの活性レベル(ニューロンの状態の内的測定値)である。

【0006】

【数2】

$$\sum_j W_{ij} \cdot X_j$$

40

【0007】

T_i はニューロンの閾値、 f_i は活性化関数(これは通常非線形である。)、 Y_i は*i* 次のニューロンの出力である。活性レベルが閾値を越えるとニューロンにより出力が生成される。

【0008】

所定の入力と関連する荷重は正であってもよく、この場合にはその入力を受信された信号は活性レベルを増加させる。従って、正の荷重は興奮性の入力とみなすことができる。ある場合には、所定の入力と関連する荷重は負であってもよく、この場合にはその入力を受信された信号は活性レベルを減少させる。従って、負の荷重は抑制性の入力とみなすこと

50

ができる。

【 0 0 0 9 】

ニューロン間の接続は荷重の値を調整することにより強められ、あるいは弱められる。例えば、所定のニューロンの特定の入力と関連する荷重がこの入力で信号が受信される度に減少してもよい。繰り返し信号（すなわち、同じ入力で複数回受信される信号）が、その入力と関連する荷重を漸次増加させてもよい。次にその入力で信号が受信されるとニューロンの活性レベルが一層増加する。

【 0 0 1 0 】

活性化関数は通常すべてのニューロンについて同一であって、固定されており、シグモイド関数（sigmoid function）が頻繁に使用される。

10

【 0 0 1 1 】

既知のニューラルネットワークにおける i 次のニューロンの活性度は、荷重 W_{ij} の値により決定される入力 x_j の単調関数（monotonic function）に限定されている。この限定により、既知のニューラルネットワークはより複雑で非単調の挙動をエミュレートする能力が制限されている。

【 0 0 1 2 】

（発明の開示）

（発明が解決しようとする技術的課題）

本発明の目的は、前述の不都合を解消するニューラルネットワークの素子を提供することにある。

20

【 0 0 1 3 】

（その解決方法）

従って、第1の発明は、複数の入力と、少なくとも一つのプロセッシング要素と、少なくとも一つの出力と、それぞれ前記少なくとも一つのプロセッシング要素に対応するアドレスに値を記憶するデジタルメモリとを備える、ニューラルネットワーク素子であって、前記少なくとも一つのプロセッシング要素は、入力信号に応答して前記デジタルメモリからの値を受信するように構成され、かつ前記デジタルメモリから受信された値に応じて複数の操作の一つを実行するように命令される、ニューラルネットワーク素子を提供する。

【 0 0 1 4 】

プロセッシング要素により受信された値が常にプロセッシング要素の活性レベルに加算される（値が常に荷重である）先行技術のニューラルネットワークとは対照的である。本発明のプロセッシング要素は、複数の命令を実行可能であり、従って、受信された値はプロセッシング要素に命令のうち実行するものを選択させる命令として機能する。

30

【 0 0 1 5 】

プロセッシング要素の活性レベルはプロセッシング要素の興奮度を示す内的値である。

【 0 0 1 6 】

好適には、前記操作は、前記少なくとも一つのプロセッシング要素の活性レベルへの前記値の加算である。

【 0 0 1 7 】

好適には、前記操作は、少なくとも一つのプロセッシング要素の活性レベルの零への減衰であり、この減衰は前記活性レベルが負であれば前記活性レベルを増加させ、前記活性レベルが正であれば前記活性レベルを減少させることによるものである。

40

【 0 0 1 8 】

好適には、前記減衰の速度は前記デジタルメモリから受信された値に依存する。

【 0 0 1 9 】

好適には、前記減衰はクロックにより制御される。

【 0 0 2 0 】

好適には、前記操作は、前記少なくとも一つのプロセッシング要素による出力の生成と、その少なくとも一つのプロセッシング要素の活性レベルの予め定められた最小レベルへのリセットとを含む。

50

【 0 0 2 1 】

好適には、前記操作は、前記少なくとも一つのプロセッシング要素により出力を生成することなく、その少なくとも一つのプロセッシング要素の活性レベルをリセットすることを含む。

【 0 0 2 2 】

好適には、前記操作は、自動学習アルゴリズムに基づいて、前記少なくとも一つの活性レベルを前記デジタルメモリに保持された値を変更することを含む。

【 0 0 2 3 】

好適には、前記操作は、自動学習アルゴリズムに基づいて、前記デジタルメモリに保持された値を変更することを含む。

10

【 0 0 2 4 】

好適には、前記自動学習アルゴリズムは、ヘブ学習アルゴリズムを備える。

【 0 0 2 5 】

好適には、前記ニューラルネットワーク素子により生成される出力は、前記出力を発生した前記プロセッシング要素の個々のアドレスの番号を含む。

【 0 0 2 6 】

好適には、前記ニューラルネットワーク素子は、前記プロセッシング要素から受信された信号パルスの時系列順序を決定し、各信号パルスを番号により表すように構成された調停及びエンコード手段を備える。例えば、各ニューロン出力がそれ自体のピンを有していれば、256本のピンが8本のピンに置換され（この場合、8ビットの2進数が使用される）、外部通信（off-chip communication）についてのピンの制限の問題が解決されるので、この構成は有利である。

20

【 0 0 2 7 】

好適には、ニューラルネットワーク素子は、前記デジタルメモリに対する読出線アクセスを有する通常のプロセッサをさらに備える。

【 0 0 2 8 】

好適には、前記通常のプロセッサによる前記デジタルメモリへのアクセスは、前記ニューラルプロセッシング要素により要求されるアクセスに対して非同期で調停される。

【 0 0 2 9 】

好適には、多数のニューラルネットワーク素子が通常のプロセッサの並列ネットワークにより管理される。

30

【 0 0 3 0 】

前記デジタルメモリはランダムアクセスメモリであることが好ましい。

【 0 0 3 1 】

好適には、前記素子は非同期法で作動する。あるいは、ニューラルネットワーク素子はクロックで作動してもよい。

【 0 0 3 2 】

第2の発明は、少なくとも一つのプロセッシング要素にそれぞれ対応するデジタルメモリのアドレスに値を記憶し、前記少なくとも一つのプロセッシング要素に値を伝送し、かつ前記デジタルメモリから受信された値に応じて前記プロセッシング要素で複数の操作のうちの一つを実行する、ニューラルネットワーク素子を操作する方法を提供する。

40

【 0 0 3 3 】

この方法は、前述の第1の発明における好適ないしは好ましい特徴を含むものであってもよい。

【 0 0 3 4 】

（発明を実施するための最良の形態）

添付図面を参照して本発明の特定の実施形態を単なる例として説明する。

【 0 0 3 5 】

図1に示すニューラルネットワークは、1024個の入力（input）1、ランダムアクセスメモリ（RAM）2、256個のプロセッシング要素3、及び256個の出力（output）

50

4を備えている。

【0036】

各入力は先行するニューラルネットワーク素子から信号を搬送する。各入力は1ビットの情報のみを搬送する。換言すれば、各入力は「オン」又は「オフ」である。入力信号は入力に対するパルスの形態を取る。

【0037】

RAM2は値の組を格納している。個々の入力及び出力の組み合わせに、RAMの異なるアドレスが割り当てられている(アドレスの個数は 1024×256 である。)。図1では、読み手の補助と図示の目的のみで、RAMを列及び行を備える二次元のマトリクスとみなすことができ、各行は特定の入力に属する値のすべてを含み、各列は特定のプロセッシング要素に属する値のすべてを含んでいる。後に詳述するように、RAMのアドレスのうちのかなりの部分が零の値を含んでいてもよい。

【0038】

所定の入力で受信されたパルスに応じて、RAM2はその入力に対応する行に配置された値を検索する。各値はその値が配置されている列と対応するプロセッシング要素3へ送られる。プロセッシング要素3の活性レベル(プロセッシング要素3の状態の内的測定値)は、受信された値に応じて変更される。各プロセッシング要素3はニューロンとして機能し、1024個の入力と1個の出力を有している。

【0039】

先行技術と異なり、RAM2に記憶される値は荷重であってもよく、命令(instruction)であってもよい。プロセッシング要素3に送られた値が命令である場合、後に詳述するように、プロセッシング要素3は命令に応じた操作を実行する。プロセッシング要素3に送られた値が荷重である場合、その荷重はプロセッシング要素3の活性レベルに加算される。「i」のプロセッシングユニットの操作は、数学的に以下のように表すことができる。

【0040】

【数3】

$$Y_i = f_i(\sum_j W_{ij} \cdot X_j - T_i)$$

【0041】

ここで X_j はニューロンに対する入力、 W_{ij} は荷重、下記の総和はニューロンの活性レベルである。

【0042】

【数4】

$$\sum_j W_{ij} \cdot X_j$$

【0043】

T_i はニューロンの閾値、 f_i は活性関数(これは通常非線形である。)、 Y_i はi次のニューロンの出力である。活性レベルが閾値を越えるとニューロンにより出力が生成される(これは事実上ニューロンの発火(firing)である。)。

【0044】

本実施形態では、特定の入力を示すjは1から1024であり、特定の出力を示すiは1から256の間である。

【0045】

多くの先行技術のニューラルネットワークは、アナログ出力値を生成するニューロンを備えている。例えば、前記の等式において、ニューロンの出力 Y_i は0と1の間の値を有する数であった。この出力が後続するニューロンに入力されると荷重が乗算され、この乗算の結果がその後続するニューロンの活性レベルに加算される。

【0046】

説明している本発明の実施形態は、アナログのニューラルネットワークとは異なり、プロ

10

20

30

40

50

セッシング要素（ニューロン）は２種類の出力、すなわち１又は０（ないしは「オン」又は「オフ」と等価）の出力のみが可能な態様で機能する。プロセッシング要素の活性レベルがその閾値を越える度にプロセッシング要素により出力パルスが生成される。外部の観察者から見れば、プロセッシング要素のパルス出力は、アナログ出力と等価とみなし得る。例えば、あるプロセッシングユニットが１秒間に２３個のパルスを生成したとすると、これは０．２３の概念的なアナログ値（notional analogue value）と等価とみなすことができる。以下、プロセッシング要素によるパルスの生成を「パルス速度エンコーディング（pulse rate encoding）」と呼ぶ。

【００４７】

一見すると、アナログ信号を使用する場合と比較すると、パルス速度エンコーディングは計算上非常に集約的であるように認められる。例えば、ニューロンの所定の入力に割り当てられた荷重に１回の計算で０．２３のアナログ信号を乗算し、その後ニューロンの活性レベルに加算することができる。これとは対照的に、パルス速度エンコーディングを使用すると、荷重をニューロンの活性レベルに対して２３回（すなわちパルス毎に）加算しなければならない、２３回の独立した計算が必要となる。しかしながら、所定のニューロンからの出力がないことが頻繁にある。パルス速度エンコーディングを使用すると、この場合には入力パルスはまったく受信されない、計算はまったく実行されない。これとは対照的に、公知のアナログニューラルネットワークでは、出力の零の値は、偶然に零となった有効な出力値である。ニューロンの荷重に零が乗算され、その積（これも零である。）がニューロンの活性レベルに加算される。このニューロンの活性レベルにまったく影響しない計算は頻繁に現れ、計算上非常に無駄である。パルス速度エンコーディングを使用すれば、この無駄な零の乗算を回避することができる。

【００４８】

ニューラルネットワーク素子は、２５６個のプロセッシング要素３を備えている。各プロセッシング要素３はニューロンとして可能し、１０２４個の入力と１個の出力を備えている。各プロセッシング要素と後続する各ニューラルネットワーク素子との間に個別の接続が設けられているとすると、接続の個数は急激に非実用的な程に多数となる。この問題を回避するため、図２に示すように、プロセッシング要素からの出力は、すべて単一の８ビットの出力ライン５に送られている。所定のプロセッシング要素からの出力パルスは、そのプロセッシング要素を代表する８ビットの２進数に変換され、この８ビットの２進数は出力ラインにより搬送される。この変換は、「アドレス・イベントエンコーディング（addressed-event encoding）」と呼ぶことができ、調停及びエンコーディングユニット（arbitrate and encoding unit）６により実行される。

【００４９】

調停及びエンコーディングユニット６の操作は図３に概略的に図示されている。このユニット６は２つの機能を備えている。すなわち、ユニット６は出力ライン５に送られるパルスの時系列順序（chronological order）を決定し、各出力をその出力を生成したプロセッシング要素を示す数字に変換する。これらの２つの機能はユニット６により同時に達成される。図３のユニット６は８個のプロセッシング要素の組について示している。

【００５０】

ユニット６の第１層は、４個の調停及びエンコーディングブロック１０～１３を備えている。第１の調停及びエンコーディングブロック１０の入力側は、２個の入力 a 及び b と、２個の確認応答出力（acknowledgement output） a_{ack} 及び b_{ack} とを備えている。プロセッシング要素の出力がハイとなると、それが入力 a で受信され、ブロック１０の第１の出力 y がハイとなり、第２の出力 z が同時に入力 a を示す１ビットの数（この場合、２進数は「１」である。）を出力する。次に、確認応答出力 a_{ack} がハイとなり、それによってプロセッシング要素に対してブロック１０により出力が処理されたことが示される。その後、プロセッシング要素の出力がローとなる。

【００５１】

入力 a と入力 b がほぼ同時にハイとなったとすると、いずれの入力が最初にハイとなった

10

20

30

40

50

かをザイツ相互排除要素 (Seitz mutual exclusion element) が決定する (相互排除要素はアディソン・ウェスレイ (Addison Wesley) から出版されたメッド (Mead) 及びコンウェイ (Conway) の「VLSIシステム入門 (Introduction to VLSI system)」の260頁等に記載されている。)。これによってブロック10からの出力が生成される。出力が生成され、確認応答が適切なプロセッシング要素に返信されると、入力のうちの後者がブロック10により処理及び出力される。

【0052】

ユニット6の第2層は一对の調停及びエンコーディングブロック14, 15を備えている。この対のうち第1のブロック14を参照すると、2個の入力 y, w が先行するブロック10, 11の出力に接続されている。ブロック14の第1の出力 u は入力信号に応じてハイとなり、同時に第2の出力 v が2ビットの数を出力する。2ビットの数の最上位の桁が入力 y を表し、2ビットの数の最下位の桁が先行するブロック10の入力 a を表している。よって、2ビットの数は2進数の「11」である。

10

【0053】

ユニット6の最終の第3層は先行する層と同様に機能する。最終の層を含む単一のブロック16からの出力は、3ビットの2進数であり、この例では数「111」である。

【0054】

図1の説明では、先行する要素3からの出力はパルスの形態で述べている。図3の説明では、プロセッシング要素からの出力は、確認応答信号がプロセッシング要素により受信された後にのみ終了する連続的なハイとして述べている。調停及びエンコーディングユニット6を正確に機能させるためには、この変更が必要となる。

20

【0055】

図2では、8ビットの2進数の出力を生成するために、8層の調停及びエンコーディングブロックが必要となる。

【0056】

図2を参照すると、4本の8ビットの入力ライン19がプロセッシング要素3及びRAM2に接続されている。調停及びエンコーディングユニット20が4本の入力ライン19間を調停するために使用されている。調停及びエンコーディングユニット20は図2に概略的に示しかつ上述したような態様で機能する。この例では、調停及びエンコーディングユニットは2層のブロックを含んでいる。調停及びエンコーディングユニットは、10ビットの出力を備え追加の2ビットは、4本の8ビットの入力ライン19間の識別のために使用されている。

30

【0057】

調停及びエンコーディングユニット20の10ビットの出力はデコーダ21に接続されている。デコーダは個々の受信した10ビットの入力数をそれぞれ対応する1ビットの出力ライン1の出力に変換する (従って、出力ラインは1024本ある。)。出力ライン1は図1に示す出力ラインに対応している。

【0058】

本発明の実施形態で使用されているパルス速度エンコーディングシステムでは、ニューラルネットワーク素子間の通信は1 μ sのタイムスケールで行われる。これは1 kHz以下の速度で発火する現在のニューロンの時間解像度と比較すると瞬間的である。

40

【0059】

ニューラルネットワーク素子がデジタルであるため、アナログのニューラルネットワーク素子にはない自由度が得られる。詳細には、RAM2に広範囲の値を記憶することができ、これによっていくつかの値をプロセッシング要素3に対する「命令」として使用することができる。この「命令」という用語は、プロセッシング要素が単に活性レベルに値を加算するのではなく操作の実行を命令されるということを指している。既知である先行技術のニューラルネットワークでは、ニューロンにより実行される機能は常に同じであり、活性レベルに値が加算される。命令を使用することにより、プロセッシング要素3はそれらの操作において大幅に柔軟性が向上する。

50

【0060】

命令を使用することにより、生物学的実際のニューロンにより近似する特性を有するニューラルネットワークが得られる。この特性の一つは、「漏洩性のある (leaky)」統合化である。換言すれば、ある期間に入力される活性度 (がなければ、活性レベルが減衰するという特性である。他の特性は「屈折性 (refraction)」であり、プロセッシング要素 3 は発火後のある期間はすべての入力を無視する。この特性によりプロセッシング要素の最大発火速度 (すなわち、プロセッシング要素 3 が出力パルスを生じ可能な速度) が制限され、それによってニューロンの活性度が絶えず増加することに起因してネットワークが不安定になることが防止される。屈折性及び漏洩性のある統合化は、ある程度リアルタイムに依存するということを包含している。

10

【0061】

屈折性及び漏洩性のある統合化は、リアルタイム参照イベント (real-time reference event)、例えば 32 kHz のクロック (図 1 及び図 2 には図示せず。) を使用することにより、説明した本発明の実施形態において実現される。クロックからのパルスは 8 ビットの入力を介して特定の 2 進数 (例えば 00000001) として伝送される。漏洩性のある統合化が必要であれば、この入力された数に対応する RAM の各アドレスに記憶された値は、プロセッシング要素に対して活性レベルが正であれば活性レベルを減少させることを命令し、活性レベルが負であれば活性レベルを増加させることを命令する。RAM に保持された値を異ならせることで、漏洩性のある統合化の速度を異ならせること、例えば、各クロックパルスに応じた減少 / 増加や、2 個のクロックパルス毎の減少 / 増加を命令し

20

【0062】

命令は自動学習アルゴリズム (例えばヘブの学習アルゴリズム (Hebbian learning algorithms)) により調整すべき否かの決定を与えるものであってもよい。自動学習能力があるニューロンの例として、以下のニューロン挙動を図 4 に示すようにモデル化することができる。

【0063】

プロセッシング要素の活性レベルは $\{-L, +T\}$ の範囲で機能し、 $-L$ は不応答レベル (refractory level)、 T は閾値である。閾値 T に達するとプロセッシング要素がパルスを出し、活性レベルは $-L$ にリセットされる。 $\{-L, 0\}$ はプロセッシング要素の不応答範囲であり、入力パルスに応じてプロセッシング要素に荷重が送られると、荷重は活性レベルに加算されず (活性レベルは変化しない。)、荷重自体は減少される。 $\{0, F\}$ は漏洩性の興奮範囲 (excited range) であり、入力パルスに応じてプロセッシング要素に荷重が送られると、荷重が活性レベルに加算され、学習法により荷重自体は変化しない。 $\{F, T\}$ は強く刺激される範囲であり、入力パルスに応じて荷重がプロセッシング要素に送られると、荷重が活性レベルに加算され、荷重自体が増加される。

30

【0064】

図 3 に図示されたモデルに従って動作するように構成されたプロセッシングユニットの動作は図 5 a から図 5 c のフローチャートに概略的に図示されている。

【0065】

最初に図 5 a を参照すると、所定の入力 j でパルスが受信される。所定の要素についてこの入力 j と関連する値は、RAM から取得され、プロセッシング要素に送られる。値は 3 つの範囲のうちの一つに入る。3 つの範囲には、活性レベルに加算される荷重と対応する第 1 の範囲と、活性レベルの加算及びプロセッシング要素に対するこの入力と関連する荷重の増加の命令に対応する第 2 の範囲と、荷重以外の命令を含む第 3 の範囲とがある。

40

【0066】

図 5 a の残りの部分は、第 1 の範囲に入る値を扱っている。値はプロセッシング要素の活性レベルに加算される (荷重が負であれば活性レベルが減少される。)。値の加算後の活性レベルが閾値 T より小さければ、それ以上の操作はなされない。活性レベルが閾値 T より大きければ、プロセッシング要素はパルスを出し、プロセッシング要素の活性レベル

50

は - L まで減少される。

【 0 0 6 7 】

値が第 2 の範囲に入れば、フローチャートに図示されたプロセッシング要素による操作は図 5 b に分岐する。プロセッシング要素の活性レベルが 0 以上であれば、値はプロセッシング要素の活性レベルに加算される。これに続いて、活性レベルが F よりも大きければ、R A M に保存された値が増加される。

【 0 0 6 8 】

プロセッシング要素の活性レベルがゼロ未満であれば、値は活性レベルに加算されず、R A M に保存された値が減少される。

【 0 0 6 9 】

値が第 3 の範囲に入れば、他の形態の命令であることを示している。例えば、図 5 c に示すように、プロセッシング要素が活性レベルとかかわりなくパルスを発生し、活性レベルを不応答レベルにリセットするように命令してもよい。

【 0 0 7 0 】

適切な値が選択されればニューラルネットワークが安定するヘブの学習アルゴリズムは、数的な荷重の処理から独立して活性レベルを - L から 0 へ戻す手順によるものであり、従って、上述のように実行される「期間にわたる零に向けての増加」の命令の使用によるものである。

【 0 0 7 1 】

プロセッシング要素が R A M の値メモリに記憶された値により規定される他の命令を実行可能であってもよいことは、当業者の理解するところである。例えば、ニューロンの発火を行うことなく、活性レベルを不応答レベルにリセットしてもよい。これは強度の抑制性入力を模している。

【 0 0 7 2 】

図 1 を参照すると、ニューラルネットワーク素子は、R A M 値記憶部 (RAM value store) 4 に対する読出線アクセス (read-wire access) を備える通常のプロセッサ (図示せず。) を含んでいてもよい。また、プロセッサは、ニューラルネットワーク素子に組み込まれたインターフェース周辺機器を使用して、ニューロン活性度ストリームに対する入力及び出力を監視及び生成可能であってもよい。

【 0 0 7 3 】

プロセッサが R A M 値記憶部 4 にアクセスし、ネットワークがその時点で飽和に近い状態で作動していなければ、ニューラルプロセッシングに対する影響を最低限としつつ、ニューロン活性度ストリームが完全に機能しているネットワークに対して非同期で調停される。一つの監視プロセッサが多数のニューラルネットワーク素子を管理してもよいが、非常に大型のシステムではこのようなプロセッサが数個必要となる。1 ギガビットの D R A M 技術に基づいて実行する場合には、各チップが監視プロセッサを有する。よって、ニューラルネットワークは、通常のプロセッサの並列ネットワークにより管理され、非常に多数のニューラルネットワークとして実行してもよい。アーキテクチャが通常のマシンに容易に接続されることは、本発明の利点である。制御プロセッサはアーキテクチャの柔軟性を大幅に拡張し、ニューラルプロセッシングと通常のシーケンシャルマシンとの間の閉じた結合 (close coupling) をもたらす。

【 0 0 7 4 】

通信がクロック化されたロジックを使用して実行されれば、時間発火情報 (temporal firing information) が量子化される。非同期のロジックが使用すれば、この量子化は回避される。従って、非同期のロジックを使用すると有利である。

【 0 0 7 5 】

図示の実施形態では 1 個のプロセッシング要素が 1 個のニューロンに対応するが、プロセッシング要素が複数のニューロンの計算を実行する役割を有していてもよいことは明らかである。

【 図面の簡単な説明 】

10

20

30

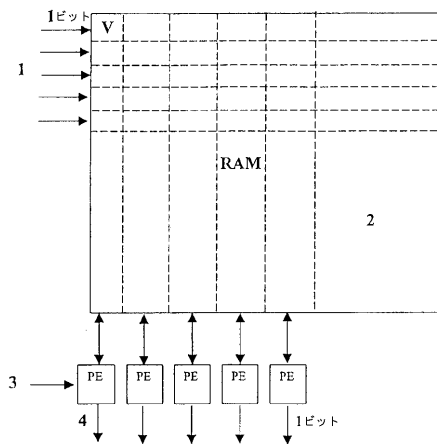
40

50

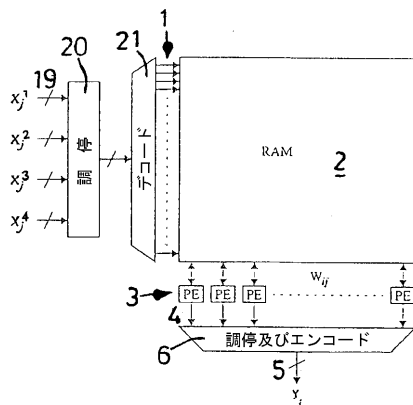
- 【図 1】 本発明に係るニューラルネットワーク素子の実施形態の概略図である。
- 【図 2】 調停及びエンコード手段を含むニューラルネットワーク素子の概略図である。
- 【図 3】 図 2 に図示された調停及びエンコード手段の概略図である。
- 【図 4】 ヘブの学習則を示すグラフである。
- 【図 5 a】 ニューラルネットワーク素子の処理を示すフローチャートである。
- 【図 5 b】 ニューラルネットワーク素子の処理を示すフローチャートである。
- 【図 5 c】 ニューラルネットワーク素子の処理を示すフローチャートである。
- 【符号の説明】
- 1 入力
 - 2 ランダムアクセスメモリ (R A M)
 - 3 プロセッシング要素
 - 4 出力
 - 5 出力ライン
 - 6 , 2 0 調停及びエンコーディングユニット
 - 1 9 入力ライン
 - 2 1 デコーダ

10

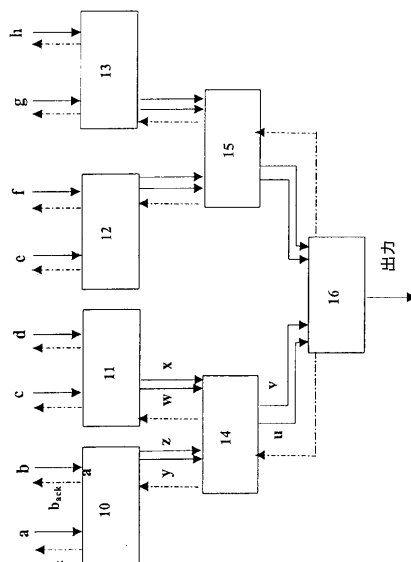
【図 1】



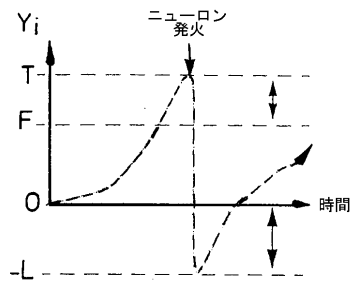
【図 2】



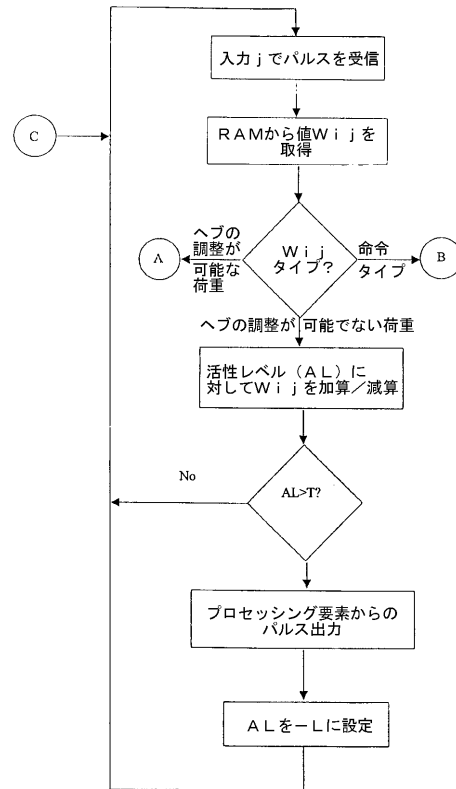
【図 3】



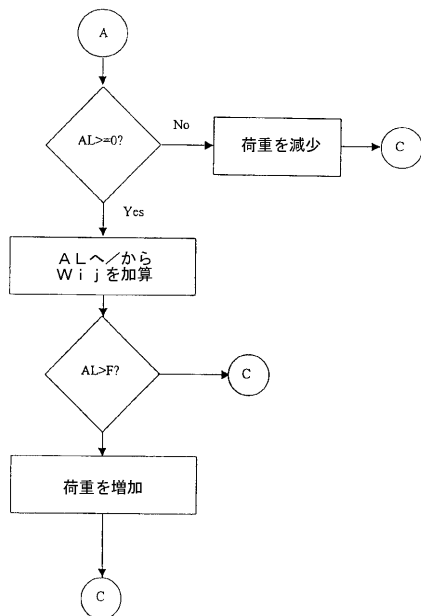
【図 4】



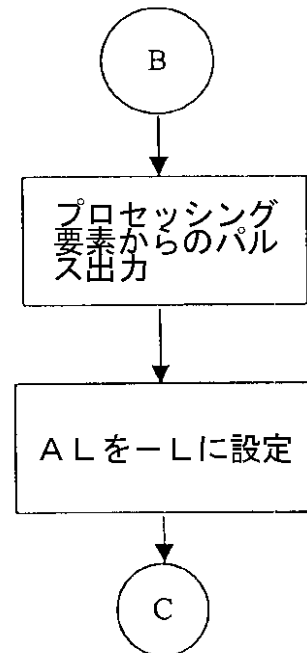
【図 5 a】



【図 5 b】



【図 5 c】



フロントページの続き

(72)発明者 スティーブン・バイルン・ファーバー
イギリス、エスケイ 9・5 ディユー、チェシャー、ウィルムズロー、ゴーシー・ロード 1 エイ番

審査官 長谷川 篤男

(56)参考文献 特開平 0 2 - 3 1 0 6 6 6 (J P , A)
特開平 0 2 - 2 3 6 6 5 8 (J P , A)
米国特許第 0 5 2 7 8 9 4 5 (U S , A)
特開平 1 0 - 0 9 1 6 0 5 (J P , A)
近藤 利夫 他, ニューラルネット情報処理技術, N T T R & D , 日本, 社団法人電気通信協会, 1 9 9 3 年 6 月 1 0 日, 第 4 2 巻 第 6 号, 第 7 7 9 - 7 8 6 頁

(58)調査した分野(Int.Cl. , D B 名)
G06N 3/063
JSTPlus(JDreamII)