

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-211972
(P2017-211972A)

(43) 公開日 平成29年11月30日(2017.11.30)

(51) Int.Cl.	F I	テーマコード (参考)
GO6N 3/04 (2006.01)	GO6N 3/04	
GO6N 3/063 (2006.01)	GO6N 3/063	

審査請求 未請求 請求項の数 8 O L (全 42 頁)

<p>(21) 出願番号 特願2016-223504 (P2016-223504)</p> <p>(22) 出願日 平成28年11月16日 (2016.11.16)</p> <p>(62) 分割の表示 特願2016-222275 (P2016-222275) の分割</p> <p>原出願日 平成28年11月15日 (2016.11.15)</p> <p>(31) 優先権主張番号 特願2016-100694 (P2016-100694)</p> <p>(32) 優先日 平成28年5月19日 (2016.5.19)</p> <p>(33) 優先権主張国 日本国(JP)</p> <p>(出願人による申告) 平成27年度 国立研究開発法人 科学技術振興機構、戦略的創造研究推進事業 (ACCL) 「TCI 積層情報処理アクセラレータの研究開発」に係る委託事業、産業技術力強化法第19条の適用を受ける特許出願</p>	<p>(71) 出願人 504173471 国立大学法人北海道大学 北海道札幌市北区北8条西5丁目</p> <p>(74) 代理人 110000958 特許業務法人 インテクト国際特許事務所</p> <p>(74) 代理人 100120189 弁理士 奥 和幸</p> <p>(74) 代理人 100173510 弁理士 美川 公司</p> <p>(72) 発明者 本村 真人 北海道札幌市北区北8条西5丁目 国立大学法人北海道大学内</p>
---	---

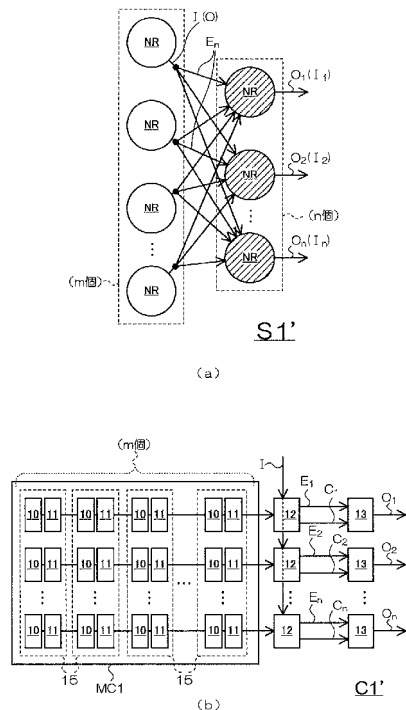
(54) 【発明の名称】 ニューラルネットワーク回路及びニューラルネットワーク集積回路

(57) 【要約】

【課題】回路規模及びコストを大幅に縮小しつつ、効率的且つ多様性を有するニューラルネットワーク回路を提供する。

【解決手段】脳機能に対応した乗算機能及び加算/適用機能を実現するニューラルネットワーク回路C1'において、重み付け係数Wとして「1」又は「0」を記憶するメモリセル10と、脳機能に対応した既定の所定値を記憶するメモリセル11と、入力データIと重み付け係数Wとの排他的論理和を生成する多数判定入力回路12と、排他的論理和と上記所定値とに基づき、当該所定値が脳機能に対応した既定値である場合に、「1」である乗算結果の総数 - 「0」である乗算結果の総数 既定の閾値) のとき「1」を出力し、(「1」である乗算結果の総数 - 「0」である乗算結果の総数 < 既定の閾値) のとき「0」を出力する直列多数判定回路13と、を備える。

【選択図】 図9



【特許請求の範囲】**【請求項 1】**

人の脳機能に対応した重み付け係数と1ビットの入力データとを乗算する乗算機能と、当該乗算機能による複数の乗算結果を加算すると共に当該加算結果に活性化関数を適用して1ビットの出力データを出力し且つ前記脳機能に対応した加算/適用機能と、を実現するニューラルネットワーク回路において、

前記重み付け係数として「1」又は「0」を記憶して第1演算部に出力する第1記憶部と、

前記脳機能に対応して予め設定された所定値を記憶して前記第1演算部に出力する第2記憶部と、

前記入力データと前記記憶されている重み付け係数との排他的論理和を前記乗算結果として生成する前記第1演算部と、

前記生成された乗算結果と、前記記憶されている所定値と、に基づいて前記加算/適用機能を実現する第2演算部であって、当該所定値が前記脳機能に対応して予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が予め定められた閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力する第2演算部と、

を備えることを特徴とするニューラルネットワーク回路。

【請求項 2】

請求項1に記載のニューラルネットワーク回路において、

前記加算/適用機能における加算機能として、前記脳機能に基づいて予め設定され且つ当該脳機能としての傾向をニューロンごとに示す整数である傾向係数であって、絶対値が0以上且つ前記入力データの数以下である傾向係数が、各前記乗算結果の加算結果に対して更に加算されるべき場合、当該傾向係数の絶対値に等しい数のいずれかの前記第1演算部に当該傾向係数が分けて入力されており、

当該傾向係数が分けて入力されている前記第1演算部は、前記所定値に応じて、各前記乗算結果の加算結果に対する当該傾向係数の更なる加算結果を前記第2演算部に出力し、

前記第2演算部は、前記傾向係数の更なる加算結果を各前記第1演算部について加算した合計値が正又は0であるとき「1」を前記出力データとして出力し、前記合計値が負であるとき「0」を前記出力データとして出力することを特徴とするニューラルネットワーク回路。

【請求項 3】

請求項1に記載のニューラルネットワーク回路において、

前記入力データの数が一であり、

前記出力データ、前記第1記憶部、前記第2記憶部、前記第1演算部及び前記第2演算部それぞれの数が、前記ニューラルネットワーク回路が対応すべき前記脳機能に基づいて予め設定された二以上の数であり、

前記重み付け係数及び前記所定値のそれぞれが前記脳機能に対応した値であり、

各前記第1演算部は、各前記出力データにそれぞれ対応する前記重み付け係数と前記入力データとの前記排他的論理和を前記乗算結果として前記出力データごとにそれぞれ生成し、

各前記第2演算部は、

前記所定値が前記予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が前記閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力することを特徴とするニューラルネットワーク回路。

【請求項 4】

請求項1に記載のニューラルネットワーク回路において、

10

20

30

40

50

前記出力データ及び前記第 2 演算部の数がそれぞれ一であり、

前記入力データ、前記第 1 記憶部、前記第 2 記憶部及び前記第 1 演算部それぞれの数が、前記ニューラルネットワーク回路が対応すべき前記脳機能に基づいて予め設定された二以上の数であり、

前記重み付け係数及び前記所定値のそれぞれが前記脳機能に対応した値であり、

各前記第 1 演算部は、各前記入力データにそれぞれ対応する前記重み付け係数と当該各入力データとの前記排他的論理和を前記乗算結果として前記入力データごとにそれぞれ生成し、

前記第 2 演算部は、

前記所定値が前記予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が前記閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力することを特徴とするニューラルネットワーク回路。

10

【請求項 5】

請求項 3 に記載のニューラルネットワーク回路と、

請求項 4 に記載のニューラルネットワーク回路と、

を備え、

一の入力データが入力される請求項 3 に記載の前記ニューラルネットワーク回路から出力される各出力データがそれぞれ、請求項 4 に記載の前記ニューラルネットワーク回路における各入力データとされており、

20

請求項 4 に記載の前記ニューラルネットワーク回路から一の出力データを出力することを特徴とするニューラルネットワーク集積回路。

【請求項 6】

請求項 5 に記載のニューラルネットワーク集積回路において、

請求項 3 に記載のニューラルネットワーク回路と請求項 4 に記載の前記ニューラルネットワーク回路との間にレジスタ部を更に備え、

当該レジスタ部は、請求項 3 に記載のニューラルネットワーク回路から出力される各出力データを一基準クロック分だけバッファリングし、請求項 4 に記載の前記ニューラルネットワーク回路における各入力データとして請求項 4 に記載の当該ニューラルネットワーク回路に出力することを特徴とするニューラルネットワーク集積回路。

30

【請求項 7】

複数の請求項 5 に記載のニューラルネットワーク集積回路が直列に接続されており、

一の当該ニューラルネットワーク集積回路のからの一の前記出力データが、直後段の当該ニューラルネットワーク集積回路への一の前記入力データとされていることを特徴とするニューラルネットワーク集積回路。

【請求項 8】

複数の請求項 5 に記載のニューラルネットワーク集積回路を備え、

当該複数の請求項 5 に記載のニューラルネットワーク集積回路のそれぞれに対して一の前記入力データが共通に入力されており、

40

当該複数の請求項 5 に記載のニューラルネットワーク集積回路からそれぞれ別個に前記出力データがずつ出力されることを特徴とするニューラルネットワーク集積回路。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ニューラルネットワーク回路及びニューラルネットワーク集積回路の技術分野に属する。より詳細には、複数の入力データと重み付け係数とをそれぞれ乗算する乗算機能と、各乗算結果を加算しその加算結果に活性化関数を適用して出力する加算/適用機能と、を有するニューラルネットワーク回路、及び当該ニューラルネットワーク回路を複数備えるニューラルネットワーク集積回路の技術分野に属する。

50

【背景技術】

【0002】

近年、人の脳機能をモデル化した、いわゆるニューラルネットワーク回路についての研究開発が行われている。このとき、従来のニューラルネットワーク回路としては、例えば浮動小数点又は固定小数点を使った積和演算を用いて実現される場合が多く、この場合には、例えば演算コストが大きく、処理負荷が高いという問題点があった。

【0003】

そこで近年、上記入力データ及び上記重み付け係数をそれぞれ1ビットとする、いわゆる「バイナリニューラルネットワーク回路」のアルゴリズムが提案されている。ここで、上記バイナリニューラルネットワーク回路のアルゴリズムを示す先行技術文献としては、例えば下記非特許文献1及び非特許文献2が挙げられる。

10

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】「XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks」論文、Mohammad Rastegari他、arXiv:1603.05279v2 [cs.CV、2016年4月19日 (URL: <http://arxiv.org/abs/1603.05279>)

【非特許文献2】「Binarized Neural Networks: Training Neural Networks with Weights and Activations Constrained to +1 or -1」論文、Matthieu Courbariaux他、arXiv:1602.02830v3 [cs.LG]、2016年3月17日 (URL: <http://arxiv.org/abs/1602.02830>)

20

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、上述したいずれの非特許文献においても、当該論文記載の理論を具体的にどのように実現するかについては、全く記載されていない。また、各論文記載の理論により単位演算コストが大幅に下がることを利用して並列演算を可能にしたいが、そのためのハードウェア構成も未知である。更に、各論文記載の理論を用いた多様な構造を持つニューラルネット回路をどのように実現するかについても検討する余地がある。

【0006】

そこで本発明は、上記の各問題点及び要請等に鑑みて為されたもので、その課題の一例は、上記バイナリニューラルネットワーク回路のアルゴリズムを用いて、回路規模及びそれに対応するコストを大幅に縮小しつつ、効率的且つ多様性を有するニューラルネットワーク回路、及び当該ニューラルネットワーク回路を備えるニューラルネットワーク集積回路を提供することにある。

30

【課題を解決するための手段】

【0007】

上記の課題を解決するために、請求項1に記載の発明は、人の脳機能に対応した重み付け係数と1ビットの入力データとを乗算する乗算機能と、当該乗算機能による複数の乗算結果を加算すると共に当該加算結果に活性化関数を適用して1ビットの出力データを出力し且つ前記脳機能に対応した加算/適用機能と、を実現するニューラルネットワーク回路において、前記重み付け係数として「1」又は「0」を記憶して第1演算部に出力するメモリセル等の第1記憶部と、前記脳機能に対応して予め設定された所定値を記憶して前記第1演算部に出力するメモリセル等の第2記憶部と、前記入力データと前記記憶されている重み付け係数との排他的論理和を前記乗算結果として生成する多数判定入力回路等の前記第1演算部と、前記生成された乗算結果と、前記記憶されている所定値と、に基づいて前記加算/適用機能を実現する第2演算部であって、当該所定値が前記脳機能に対応して予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が予め定められた閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力する直列多数判定

40

50

回路等の第2演算部と、を備える。

【0008】

請求項1に記載の発明によれば、人の脳機能に対応し且つ各1ビットの入力データ及び出力データがそれぞれ入出力されるニューラルネットワーク回路における乗算機能を第1記憶部、第2記憶部及び第1演算部により実現し、更に当該ニューラルネットワーク回路における加算/適用機能を第2演算部により実現するので、回路規模及びそれに対応するコストを大幅に縮小しつつ、人の脳機能に対応したニューラルネットワーク回路を効率的に実現することができる。

【0009】

上記の課題を解決するために、請求項2に記載の発明は、請求項1に記載のニューラルネットワーク回路において、前記加算/適用機能における加算機能として、前記脳機能に基づいて予め設定され且つ当該脳機能としての傾向をニューロンごとに示す整数である傾向係数であって、絶対値が0以上且つ前記入力データの数以下である傾向係数が、各前記乗算結果の加算結果に対して更に加算されるべき場合、当該傾向係数の絶対値に等しい数のいずれかの前記第1演算部に当該傾向係数が分けて入力されており、当該傾向係数が分けて入力されている前記第1演算部は、前記所定値に応じて、各前記乗算結果の加算結果に対する当該傾向係数の更なる加算結果を前記第2演算部に出力し、前記第2演算部は、前記傾向係数の更なる加算結果を各前記第1演算部について加算した合計値が正又は0であるとき「1」を前記出力データとして出力し、前記合計値が負であるとき「0」を前記出力データとして出力するように構成される。

10

20

【0010】

請求項2に記載の発明によれば、請求項1に記載の発明の作用に加えて、脳機能に基づく既定の傾向係数が各乗算結果の加算結果に対して更に加算されるべき場合、当該傾向係数の絶対値に等しい数のいずれかの第1演算部に当該傾向係数が分けて入力される。そして上記所定値に応じて、各乗算結果の加算結果に対する当該傾向係数の更なる加算結果が第1演算部から第2演算部に出力される。これらにより、傾向係数の更なる加算結果を各第1演算部について加算した合計値が正又は0であるとき「1」が出力データとして第2演算部から出力され、当該合計値が負であるとき「0」が出力データとして第2演算部から出力される。よって、脳機能に基づく既定の傾向係数が導入されるべき場合であっても、回路規模及びそれに対応するコストを大幅に縮小しつつ、人の脳機能に対応したニューラルネットワーク回路を効率的に実現することができる。

30

40

【0011】

上記の課題を解決するために、請求項3に記載の発明は、請求項1に記載のニューラルネットワーク回路において、前記入力データの数が一であり、前記出力データ、前記第1記憶部、前記第2記憶部、前記第1演算部及び前記第2演算部それぞれの数が、前記ニューラルネットワーク回路が対応すべき前記脳機能に基づいて予め設定された二以上の数であり、前記重み付け係数及び前記所定値のそれぞれが前記脳機能に対応した値であり、各前記第1演算部は、各前記出力データにそれぞれ対応する前記重み付け係数と前記入力データとの前記排他的論理和を前記乗算結果として前記出力データごとにそれぞれ生成し、各前記第2演算部は、前記所定値が前記予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が前記閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力するように構成される。

【0012】

請求項3に記載の発明によれば、請求項1に記載の発明の作用に加えて、入力データの数が一であり、出力データ、第1記憶部、第2記憶部、第1演算部及び第2演算部それぞれの数が、対応すべき脳機能に基づいた既定の二以上の数とされており、重み付け係数及び所定値のそれぞれが脳機能に対応した値とされている。そして各第1演算部が、各出力データにそれぞれ対応する重み付け係数と入力データとの排他的論理和を乗算結果として

50

出力データごとにそれぞれ生成し、各第2演算部が、「1」である乗算結果の総数 - 「0」である乗算結果の総数 閾値)であるとき「1」を出力データとして出力し、「1」である乗算結果の総数 - 「0」である乗算結果の総数 < 閾値)であるとき「0」を出力データとして出力する。よって、各第1記憶部、各第2記憶部及び各第1演算部により出力データごとに乗算機能を実現し、更に各第2演算部により出力データごとに加算/適用機能を実現するので、回路規模及びそれに対応するコストを大幅に縮小しつつ、いわゆる直列入力 - 並列出力型のニューラルネットワーク回路を効率的に実現することができる。

【0013】

上記の課題を解決するために、請求項4に記載の発明は、請求項1に記載のニューラルネットワーク回路において、前記出力データ及び前記第2演算部の数がそれぞれ一であり、前記入力データ、前記第1記憶部、前記第2記憶部及び前記第1演算部それぞれの数が、前記ニューラルネットワーク回路が対応すべき前記脳機能に基づいて予め設定された二以上の数であり、前記重み付け係数及び前記所定値のそれぞれが前記脳機能に対応した値であり、各前記第1演算部は、各前記入力データにそれぞれ対応する前記重み付け係数と当該各入力データとの前記排他的論理和を前記乗算結果として前記入力データごとにそれぞれ生成し、前記第2演算部は、前記所定値が前記予め設定された値である場合において、「1」である前記乗算結果の総数から「0」である前記乗算結果の総数を減じた値が前記閾値以上であるとき、「1」を前記出力データとして出力し、前記所定値が前記予め設定された値である場合において、前記減じた値が前記閾値未満であるとき、「0」を前記出力データとして出力するように構成される。

10

20

【0014】

請求項4に記載の発明によれば、請求項1に記載の発明の作用に加えて、出力データ及び第2演算部の数がそれぞれ一であり、入力データ、第1記憶部、第2記憶部及び第1演算部それぞれの数が、対応すべき脳機能に基づいた既定の二以上の数とされており、重み付け係数及び所定値のそれぞれが脳機能に対応した値とされている。そして各第1演算部が、各入力データにそれぞれ対応する重み付け係数と当該各入力データとの排他的論理和を乗算結果として入力データごとにそれぞれ生成し、第2演算部が、「1」である乗算結果の総数 - 「0」である乗算結果の総数 閾値)であるとき「1」を出力データとして出力し、「1」である乗算結果の総数 - 「0」である乗算結果の総数 < 閾値)であるとき「0」を出力データとして出力する。よって、各第1記憶部、各第2記憶部及び各第1演算部により入力データごとに乗算機能を実現し、更に一の第2演算部により入力データごとに加算/適用機能を実現するので、回路規模及びそれに対応するコストを大幅に縮小しつつ、いわゆる並列入力 - 直列出力型のニューラルネットワーク回路を効率的に実現することができる。

30

【0015】

上記の課題を解決するために、請求項5に記載の発明は、請求項3に記載のニューラルネットワーク回路と、請求項4に記載のニューラルネットワーク回路と、を備え、一の入力データが入力される請求項3に記載の前記ニューラルネットワーク回路から出力される各出力データがそれぞれ、請求項4に記載の前記ニューラルネットワーク回路における各入力データとされており、請求項4に記載の前記ニューラルネットワーク回路から一の出力データを出力するように構成される。

40

【0016】

請求項5に記載の発明によれば、請求項3に記載のニューラルネットワーク回路と、請求項4に記載のニューラルネットワーク回路と、を備え、一の入力データが入力される請求項3に記載のニューラルネットワーク回路から出力される各出力データがそれぞれ、請求項4に記載のニューラルネットワーク回路における各入力データとされており、請求項4に記載のニューラルネットワーク回路から一の出力データが出力される。よって、本発明に係る直列入力 - 並列出力型のニューラルネットワーク回路と本発明に係る並列入力 - 直列出力型のニューラルネットワーク回路とを接続することで、回路規模及びそれに対応するコストを大幅に縮小しつつ、多様な脳機能に対応可能なニューラルネットワーク回路

50

を効率的に実現することができる。

【0017】

上記の課題を解決するために、請求項6に記載の発明は、請求項5に記載のニューラルネットワーク集積回路において、請求項3に記載のニューラルネットワーク回路と請求項4に記載の前記ニューラルネットワーク回路との間にパイプラインレジスタ等のレジスタ部を更に備え、当該レジスタ部は、請求項3に記載のニューラルネットワーク回路から出力される各出力データを一基準クロック分だけバッファリングし、請求項4に記載の前記ニューラルネットワーク回路における各入力データとして請求項4に記載の当該ニューラルネットワーク回路に出力するように構成される。

【0018】

請求項6に記載の発明によれば、請求項5に記載の発明の作用に加えて、請求項3に記載のニューラルネットワーク回路と請求項4に記載の前記ニューラルネットワーク回路との間に備えられたレジスタ部により、請求項3に記載のニューラルネットワーク回路から出力される各出力データを一基準クロック分だけバッファリングして、請求項4に記載のニューラルネットワーク回路における各入力データとして請求項4に記載の当該ニューラルネットワーク回路に出力する。よって、請求項4に記載のニューラルネットワーク回路における第2演算部における見かけ上の並列処理が可能となり、多様な脳機能に対応可能なニューラルネットワーク回路を効率的且つ高速化して実現することができる。

【0019】

上記の課題を解決するために、請求項7に記載の発明は、複数の請求項5に記載のニューラルネットワーク集積回路が直列に接続されており、一の当該ニューラルネットワーク集積回路のからの一の前記出力データが、直後段の当該ニューラルネットワーク集積回路への一の前記入力データとされているように構成される。

【0020】

請求項7に記載の発明によれば、請求項5に記載の発明の作用に加えて、複数の請求項5に記載のニューラルネットワーク集積回路が直列に接続されており、一の当該ニューラルネットワーク集積回路のからの一の出力データが、直後段の当該ニューラルネットワーク集積回路への一の入力データとされているので、より多様な脳機能に対応可能なニューラルネットワーク回路を効率的に実現することができる。

【0021】

上記の課題を解決するために、請求項8に記載の発明は、複数の請求項5に記載のニューラルネットワーク集積回路を備え、当該複数の請求項5に記載のニューラルネットワーク集積回路のそれぞれに対して一の前記入力データが共通に入力されており、当該複数の請求項5に記載のニューラルネットワーク集積回路からそれぞれ別個に前記出力データが一つずつ出力されるように構成される。

【0022】

請求項8に記載の発明によれば、請求項5に記載の発明の作用に加えて、複数の請求項5に記載のニューラルネットワーク集積回路のそれぞれに対して一の入力データが共通に入力されており、当該複数の請求項5に記載のニューラルネットワーク集積回路からそれぞれ別個に出力データが一つずつ出力されるので、より多様な脳機能に対応可能なニューラルネットワーク回路を効率的に実現することができる。

【発明の効果】

【0023】

本発明によれば、人の脳機能に対応し且つ各1ビットの入力データ及び出力データがそれぞれ入出力されるニューラルネットワーク回路における乗算機能を第1記憶部、第2記憶部及び第1演算部により実現し、更に当該ニューラルネットワーク回路における加算/適用機能を第2演算部により実現する。

【0024】

従って、回路規模及びそれに対応するコストを大幅に縮小しつつ、人の脳機能に対応したニューラルネットワーク回路及びその集積回路を効率的に実現することができる。

10

20

30

40

50

【図面の簡単な説明】

【0025】

【図1】関連形態に係るニューラルネットワークを説明する図であり、(a)は一つのニューロンをモデル化したユニットを示す図であり、(b)は複数のユニットの結合からなるニューラルネットワークの状態を示す図である。

【図2】関連形態に係るニューラルネットワーク回路を示す図であり、(a)は当該ニューラルネットワーク回路に相当するニューラルネットワークを示す図であり、(b)は当該ニューラルネットワーク回路の構成を示すブロック図であり、(c)は当該ニューラルネットワーク回路に対応する真理値表である。

【図3】関連形態に係るニューラルネットワーク回路の細部構成を示す図であり、(a)は当該細部構成に係るメモリセルの回路の一例を示す図であり、(b)は当該細部構成の回路の一例を示す図である。

【図4】関連形態に係るニューラルネットワーク集積回路の第1例を示す図であり、(a)は当該第1例に相当するニューラルネットワークを示す図であり、(b)は当該第1例の構成を示すブロック図である。

【図5】関連形態に係るニューラルネットワーク集積回路の第2例を示す図であり、(a)は当該第2例に相当するニューラルネットワークを示す図であり、(b)は当該第2例の構成を示すブロック図である。

【図6】関連形態に係るニューラルネットワーク集積回路の第3例を示す図であり、(a)は当該第3例に相当するニューラルネットワークを示す図であり、(b)は当該第3例の構成を示すブロック図である。

【図7】関連形態に係るニューラルネットワーク集積回路の第4例を示す図であり、(a)は当該第4例に相当するニューラルネットワークを示す図であり、(b)は当該第4例の構成を示すブロック図であり、(c)は当該第4例に係るスイッチボックスの構成の一例を示すブロック図である。

【図8】実施形態に係るニューラルネットワーク集積回路の第1例の一部を示す図であり、(a)は当該一部に相当するニューラルネットワークを示す図であり、(b)は当該一部の構成を示すブロック図であり、(c)の当該一部に対応する真理値表である。

【図9】実施形態に係るニューラルネットワーク集積回路の第1例を示す図であり、(a)は当該第1例に相当するニューラルネットワークを示す図であり、(b)は当該第1例の構成を示すブロック図である。

【図10】実施形態に係るニューラルネットワーク回路の第1例を示す図であり、(a)は当該第1例に相当するニューラルネットワークを示す図であり、(b)は当該第1例の構成を示すブロック図である。

【図11】実施形態に係るニューラルネットワーク集積回路の第2例を示す図であり、(a)は当該第2例に相当するニューラルネットワークを示す図であり、(b)は当該第2例の構成を示すブロック図である。

【図12】実施形態に係るニューラルネットワーク集積回路の第3例を示す図であり、(a)は当該第3例に相当するニューラルネットワークを示す図であり、(b)は当該第3例の構成を示すブロック図である。

【図13】実施形態に係るニューラルネットワーク集積回路の第4例を示す図であり、(a)は当該第4例の構成を示すブロック図であり、(b)は当該第4例に相当する回路例を示す図である。

【図14】実施形態に係るニューラルネットワーク集積回路の第4例の細部構成を示す図であり、(a)は当該第4例に係るパイプラインレジスタ等の回路の一例を示す図であり、(b)は当該第4例に係る多数判定入力回路及び直列多数判定回路それぞれの一例を示す図であり、(c)は当該第4例に係る並列多数判定回路の一例を示す図であり、(d)は当該第4例における動作を示すタイミングチャートである。

【図15】実施形態に係るニューラルネットワーク集積回路の第5例を示す図であり、(a)は当該第5例の構成を示すブロック図であり、(b)は当該第5例に係る制御部の細

10

20

30

40

50

部構成を示すブロック図である。

【図16】実施形態に係るニューラルネットワーク集積回路の第6例等の構成を示すブロック図であり、(a)は当該第6例の構成を示すブロック図であり、(b)は実施形態に係るニューラルネットワーク集積回路の第7例の構成を示すブロック図である。

【図17】変形形態に係るニューラルネットワークを説明する図であり、(a)は変形形態に係る一つのニューロンをモデル化したユニットを示す図であり、(b)は変形形態に係る傾向変数と出力データとの関係等を示す図である。

【図18】変形形態に係るニューラルネットワーク集積回路の細部構成を示す図であり、(a)は当該細部構成に相当する回路の一例を示す図であり、(b)は当該細部構成としての多数判定入力回路及び直列多数判定回路それぞれの一例を示す図であり、(c)は当該細部構成としての並列多数判定回路の一例を示す図である。

【発明を実施するための形態】

【0026】

次に、本発明に係る関連形態及び実施形態について、図面に基づいてそれぞれ説明する。なお以下に説明する関連形態等は、人の脳機能をモデル化したニューラルネットワークを電子的な回路で実現したニューラルネットワーク回路について本発明を適用した場合の関連形態等である。

(I) 関連形態

初めに、本発明に関連する関連形態について、図1乃至図7を用いて説明する。なお、図1は関連形態に係るニューラルネットワークを説明する図であり、図2は関連形態に係るニューラルネットワーク回路を示す図であり、図3は当該ニューラルネットワーク回路の細部構成を示す図である。また、図4は関連形態に係るニューラルネットワーク集積回路の第1例を示す図であり、図5は当該ニューラルネットワーク集積回路の第2例を示す図であり、図6は当該ニューラルネットワーク集積回路の第3例を示す図であり、図7は当該ニューラルネットワーク集積回路の第4例を示す図である。

【0027】

(A) ニューラルネットワークについて

先ず、上記脳機能をモデル化したニューラルネットワークについて、一般的に図1を用いて説明する。

【0028】

一般に人の脳の中には、多数のニューロン(神経細胞)が存在しているとされている。脳の中で各ニューロンは、多数の他のニューロンからの電気信号を受信し、また更に他の多数のニューロンへ電気信号を送信している。そして脳は、各ニューロン間のこれら電気信号の流れによって、様々な情報処理を行っていると考えられている。このとき、各ニューロン間における電気信号の送受信は、シナプスと呼ばれる細胞を介して行われる。そして、脳における上記ニューロン間の電気信号の送受信をモデル化してコンピュータ内に脳機能を実現しようとしたものが、ニューラルネットワークである。

【0029】

より具体的にニューラルネットワークでは図1(a)に例示するように、外部から入力される複数の入力データ I_1 、入力データ I_2 、...、入力データ I_n (n は自然数。以下同様。)のそれぞれに対する乗算処理、加算処理及び活性化関数の適用処理を一つのニューロン NR で実行し、その結果を出力データ O とすることで、脳機能における一つのニューロンに対する上記電気信号の送受信をモデル化する。なお以下の説明において、上記活性化関数の適用処理を、単に「活性化処理」と称する。このとき一つのニューロン NR では、複数の入力データ I_1 、入力データ I_2 、...、入力データ I_n それぞれに対応して予め設定された(つまり既定の)重み付け係数 W_1 、重み付け係数 W_2 、...、重み付け係数 W_n を当該入力データ I_1 、入力データ I_2 、...、入力データ I_n に対してそれぞれ乗算することで上記乗算処理が実行される。その後当該ニューロン NR は、各入力データ I_1 、入力データ I_2 、...、入力データ I_n に対する上記乗算処理の結果のそれぞれを加算する上記加算処理を実行する。そして当該ニューロン NR は次に、上記加算処理の結果に既

10

20

30

40

50

定の活性化関数 F を適用する上記活性化処理を実行し、その結果を上記出力データ O とし、他の一又は複数のニューロンに NR に出力する。上述した一連の乗算処理、加算処理及び活性化処理を数式で表すと、図 1 (a) に示す式 (1) となる。このとき、重み付け係数 W_1 、重み付け係数 W_2 、...、重み付け係数 W_n を入力データ I_1 、入力データ I_2 、...、入力データ I_n にそれぞれ乗算する乗算処理が、ニューロン NR 間の上記電気信号のやり取りにおける上記シナプスの作用に相当する。そして図 1 (b) に例示されるように、図 1 (a) に例示する一つのニューロン NR が多数集まってシナプスにより互いに接続されることにより、脳全体がニューラルネットワーク SS としてモデル化される。なお以下の説明において、入力データ I_1 乃至入力データ I_n 又は入力データ I_m (m は自然数。以下同様。) に共通の事項を説明する場合、単に「入力データ I 」と称する。また同様に、出力データ O_1 乃至出力データ O_n 又は出力データ O_m に共通の事項を説明する場合、単に「出力データ O 」と称する。更に同様に、重み付け係数 W_1 乃至重み付け係数 W_n 又は重み付け係数 W_m に共通の事項を説明する場合、単に「重み付け係数 W 」と称する。

10

【 0 0 3 0 】

そして、以下に説明する関連形態等に係るニューラルネットワーク回路又はニューラルネットワーク集積回路は、図 1 を用いて説明した一般的なニューラルネットワークを、上記非特許文献 1 又は非特許文献 2 に記載されている手法によりバイナリ化したニューラルネットワーク回路又はニューラルネットワーク集積回路によりモデル化するものである。

【 0 0 3 1 】

(B) 関連形態に係るニューラルネットワーク回路について

20

次に、関連形態に係るニューラルネットワーク回路について、図 2 及び図 3 を用いて例示しつつ説明する。

【 0 0 3 2 】

図 2 (a) に示すように、当該ニューラルネットワーク回路に相当するニューラルネットワーク S では、一つのニューロン NR に対して例えば四つの他のニューロン NR から 1 ビットの入力データ I がそれぞれ入力され、それに対応する出力データ O が当該ニューロン NR から出力される。このとき入力データ I は、その出力元のニューロン NR から見れば 1 ビットの出力データ O となる。また 1 ビットの出力データ O は、その出力先のニューロン NR から見ると 1 ビットの入力データ I となる。上記の通り入力データ I 及び出力データ O はそれぞれ 1 ビットであるため、入力データ I の値及び出力データ O の値は、いずれも「 0 」又は「 1 」のいずれかである。そして、図 2 (a) において四つの入力データ I が入力されているニューロン NR (図 2 (a) においてハッチングで示される) において実行される上記乗算処理等に相当する上記式 (1) は、上記式 (1) において $n = 4$ とした場合の式である。即ち上記ニューラルネットワーク S は、並列多入力 - 一出力型の一段ニューラルネットワークである。

30

【 0 0 3 3 】

次に、図 2 (a) に示すニューラルネットワーク S においてハッチングで示されるニューロン NR に相当する関連形態に係るニューラルネットワーク回路の構成を、図 2 (b) にニューラルネットワーク回路 CS として示す。当該ニューラルネットワーク回路 CS は、各々が 1 ビットの入力データ I_1 乃至入力データ I_4 にそれぞれ対応する四つのメモリセル 1 と、多数判定回路 2 と、により構成される。この構成において各メモリセル 1 は、「 1 」、又は「 0 」、或いは「 NC (Not Connected (接続なし)) 」を意味する所定値の三つのいずれか一つを記憶値として記憶すると共に、比較機能を有する三値のメモリセルである。そして各メモリセル 1 は、それぞれへの入力データ I の値と、それぞれの記憶値と、に応じた値を有する出力データ E_1 乃至出力データ E_4 を、多数判定回路 2 にそれぞれ出力する。

40

【 0 0 3 4 】

ここで、メモリセル 1 の記憶値の一つである上記所定値が意味する上記「 NC 」は、関連形態に係るニューラルネットワーク S における二つのニューロン NR 間に接続がない状態である。即ち、そのメモリセル 1 が対応している二つのニューロン NR (即ち入力ニュー

50

ーロンと出力ニューロン)が接続されていない場合、そのメモリセル1の記憶値は上記所定値に設定される。一方、メモリセル1の他の記憶値(「1」又は「0」)のいずれをそのメモリセル1に記憶させておくかは、そのメモリセル1が対応している接続により接続される二つのニューロンNR間の当該接続における重み係数Wに基づいて設定される。ここで各メモリセル1にどのような記憶値を記憶させておくかは、ニューラルネットワークSとしてどのような脳機能をモデル化するか(より具体的には、例えばニューラルネットワークSを構成するニューロンNR間の接続状態等)等に基づいて予め設定されている。なお以下の説明において、出力データ E_1 乃至出力データ E_n に共通の事項を説明する場合、単に「出力データE」と称する。

【0035】

そして、各メモリセル1における上記記憶値及びそれぞれに入力される入力データIの値と、各メモリセル1から出力される出力データEの値と、の関係は、図2(c)に真理値表を示す関係とされる。即ち各メモリセル1は、当該各メモリセル1の記憶値と入力データIの値との排他的否定論理和を当該各メモリセル1から出力データEとして出力する。また各メモリセル1の記憶値が上記所定値である場合、そのメモリセル1からは、入力データIがいずれの値であっても当該所定値が出力データEとして多数判定回路2に出力される。なお、各メモリセル1の細部構成については、後ほど図3(a)を用いて説明する。

【0036】

次に多数判定回路2は、各メモリセル1からの出力データEの値に基づき、値「1」の出力データEの数が値「0」の出力データEの数より大きい場合にのみ値「1」の出力データEを出力し、それ以外の場合に値「0」の出力データEを出力する。このとき、値「1」の出力データEの数が値「0」の出力データEの数より大きい場合以外の場合とは、具体的には、いずれかのメモリセル1から値「NC」が出力されている場合、又は各メモリセル1からの値「1」の出力データEの数が値「0」の出力データEの数以下の場合、のいずれかである。なお多数判定回路2及び各メモリセル1を含むニューラルネットワーク回路CSの細部構成については、後ほど図3(b)を用いて説明する。

【0037】

ここでニューラルネットワーク回路CSは上述したように、図2(a)においてハッチングで示されるニューロンNRにおける上記乗算処理、加算処理及び活性化処理をモデル化した回路である。そして、各メモリセル1からの上記排他的否定論理和としての出力データEの出力が上記重み付け係数Wを用いた上記乗算処理に相当する。また多数判定回路2は、値「1」の出力データEの数と値「0」の出力データEの数とを比較する比較処理の前提として、値「1」の出力データEの数を加算してその合計値を算出すると共に、値「0」の出力データEの数を加算してその合計値を算出する。これらの加算が上記加算処理に相当する。そして多数判定回路2において、値「1」の出力データE及び値「0」の出力データEそれぞれの数の上記合計値を比較し、前者の数から後者の数を減じた値が予め設定された多数判定閾値以上の場合にのみ、値「1」の出力データEを多数判定回路2から出力する。一方それ以外の場合、即ち値「1」の出力データEの数の合計値から値「0」の出力データEの数の合計値を減じた値が多数判定閾値未満の場合に値「0」の出力データEを多数判定回路2から出力する。このとき、出力データEが上記所定値の場合に多数判定回路2は、当該出力データEを、値「1」の出力データEの数及び値「0」の出力データEの数のいずれにも加算しない。

【0038】

ここで、多数判定回路2における上記多数判定閾値を用いた処理について、より具体的に説明する。なお図2に例示するニューラルネットワーク回路CSでは、値「1」の出力データEの数と値「0」の出力データEの数との総数は「4」であるが、説明の明確化のため、当該総数が「10」である場合の上記処理について説明する。

【0039】

即ち、例えば多数判定閾値が「0」であり、値「1」の出力データEの数と値「0」の

10

20

30

40

50

出力データEの数が共に「5」であるとする、値「1」の出力データEの数から値「0」の出力データEの数を減じた値は「0」であり、これは当該多数判定閾値と等しい。よってこの場合に多数決判定回路2は、値「1」の出力データOを出力する。これに対して、多数判定閾値が「0」であり、値「1」の出力データEの数が「4」であり、値「0」の出力データEの数が「6」であるとする、値「1」の出力データEの数から値「0」の出力データEの数を減じた値は「-2」であり、これは当該多数判定閾値より小さい。よってこの場合に多数決判定回路2は、値「0」の出力データOを出力する。

【0040】

他方、例えば多数判定閾値が「-2」であり、値「1」の出力データEの数と値「0」の出力データEの数が共に「5」であるとする、値「1」の出力データEの数から値「0」の出力データEの数を減じた値「0」は当該多数判定閾値より大きい。よってこの場合に多数決判定回路2は、値「1」の出力データOを出力する。これに対して、多数判定閾値が「-2」であり、値「1」の出力データEの数が「4」であり、値「0」の出力データEの数が「6」であるとする、値「1」の出力データEの数から値「0」の出力データEの数を減じた値「-2」は当該多数判定閾値と等しい。よってこの場合も多数決判定回路2は、値「1」の出力データOを出力する。

10

【0041】

以上具体的に説明した多数判定回路2における処理が上記活性化処理に相当する。以上の通り、図2(b)に示すニューラルネットワーク回路CSにより、図2(a)においてハッチングで示されるニューロンNRとしての各処理がモデル化される。

20

【0042】

次に、各メモリセル1の細部構成について、図3(a)を用いて説明する。図3(a)に示すように、各メモリセル1のそれぞれは、トランジスタ T_1 乃至トランジスタ T_{14} と、インバータ IV_1 乃至インバータ IV_4 と、により構成されている。なお図3に示す各トランジスタ T_1 等のそれぞれは、例えばMOSFET(Metal Oxide semiconductor Field Effect Transistor)等により構成されている。そしてこれらの素子が、入力データ I_n に相当する接続線 LI_n 及び接続線/ LI_n 、ワード(Word)信号に相当する接続線 $W1$ 及び接続線 $W2$ 、並びにマッチ(match)信号に相当するマッチ線 M 及び反転マッチ線/ M により図3(a)に示す態様で接続されて、一つのメモリセル1が構成されている。このとき、トランジスタ T_1 及びトランジスタ T_2 、並びにインバータ IV_1 及びインバータ IV_2 により例えばSRAM(static random access memory)としての一のメモリセル 1_1 が構成され、トランジスタ T_3 及びトランジスタ T_4 、並びにインバータ IV_3 及びインバータ IV_4 により例えばSRAMとしての一のメモリセル 1_2 が構成される。また、トランジスタ T_5 乃至トランジスタ T_9 によりXNORゲート G_1 が構成され、トランジスタ T_{10} 乃至トランジスタ T_{14} によりXORゲート G_2 が構成される。

30

【0043】

次に、多数判定回路2及び各メモリセル1を含むニューラルネットワーク回路CSの細部構成について、図3(b)を用いて説明する。なお図3(b)は、図2(a)に対応して入力データIが四つである(即ち、メモリセル1を四つ備える)ニューラルネットワーク回路CSの細部構成について示している。なお図3(b)に例示するニューラルネットワーク回路CSでは、上記多数判定閾値が「0」である場合について説明する。

40

【0044】

図3(b)に示すようにニューラルネットワーク回路CSは、四つのメモリセル1と、多数判定回路2を構成するトランジスタ T_{20} 乃至トランジスタ T_{30} (図3(b)破線参照)と、により構成されている。このとき、図3(b)において一点鎖線で示されるように、トランジスタ T_{25} 乃至トランジスタ T_{28} により、フリップフロップ型のセンスアンプSAが構成されている。そしてこれらの素子が、四つのメモリセル1に共通の上記接続線 $W1$ 及び接続線 $W2$ 並びに上記接続線 M 及び接続線/ M 、及び出力データOに相当する接続線 LO 及び接続線/ LO により図3(b)に示す態様で接続されて、一つのニューラルネットワーク回路CSが構成されている。また図3(b)に示すニューラルネット

50

ワーク回路CSには、当該ニューラルネットワーク回路CSとしての処理を規定するための予め設定されたタイミング信号 T_1 、タイミング信号 T_2 及びタイミング信号 T_3 並びにタイミング信号 T_4 が外部から入力されている。このとき、タイミング信号 T_1 はトランジスタ T_{20} 乃至トランジスタ T_{22} のゲート端子にそれぞれ入力され、タイミング信号 T_2 及びタイミング信号 T_3 はトランジスタ T_{29} 及びトランジスタ T_{30} のゲート端子にそれぞれ入力され、タイミング信号 T_4 はトランジスタ T_{23} 及びトランジスタ T_{24} のゲート端子にそれぞれ入力されている。以上の構成において、タイミング信号 T_1 に基づいてそれぞれプリチャージされた各メモリセル1のマッチ線Mと反転マッチ線/Mとでは、入力データIの値並びにメモリCL $_1$ 及びメモリCL $_2$ の記憶値に応じて、当該プリチャージされた電荷が引き抜かれるタイミングが異なる。そしてセンスアンプSAは、これらマッチ線M又は反転マッチ線/Mのどちらがより早く当該プリチャージされた電荷を引き抜かれるかを検出し、更に当該マッチ線Mと反転マッチ線/Mとの間の電圧差を増幅することにより、当該検知結果を接続線LO及び接続線/LOに出力する。ここで、接続線LOにおける値が「1」であることが、ニューラルネットワーク回路CSとしての出力データOの値「1」であることを意味することになる。以上の構成及び動作によりニューラルネットワーク回路CSは、上記タイミング信号 T_1 等に基づいて、図2(a)においてハッチングで示されるニューロンNRとしての各処理をモデル化した処理を実行し、上記出力データOを出力する。

10

【0045】

(C) 関連形態に係るニューラルネットワーク集積回路の第1例について

20

次に、関連形態に係るニューラルネットワーク集積回路の第1例について、図4を用いて説明する。なお図4において、図2及び図3を用いて説明した関連形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【0046】

以下の図4乃至図7を用いてそれぞれ説明する関連形態に係るニューラルネットワーク集積回路は、図2及び図3を用いて説明した関連形態に係るニューラルネットワーク回路を複数集積した集積回路である。そしてこれらのニューラルネットワーク集積回路は、より多くのニューロンNRからなる複雑なニューラルネットワークをモデル化するためのものである。

30

【0047】

先ず、図4(a)に例示するニューラルネットワークS1をモデル化するための関連形態に係るニューラルネットワーク集積回路の第1例について説明する。当該ニューラルネットワークS1は、図4(a)においてハッチングで示されるm個のニューロンNRのそれぞれに対してn個のニューロンNRから1ビットの出力データOがそれぞれ出力されることにより、当該ハッチングで示されるニューロンNRから1ビットの出力データOがそれぞれ出力されるニューラルネットワークである。即ち上記ニューラルネットワークS1は、並列多入力・並列多出力型の一段ニューラルネットワークである。ここで図4(a)においては、各ニューロンNRの全てが入力信号I又は出力信号Oにより接続されている場合を示しているが、モデル化しようとする脳機能に応じて、各ニューロンNRのいずれかの間が接続されていなくてもよい。そしてこのことが、当該接続されていないニューロンNR間の接続に対応するメモリセル1の記憶値として上記所定値が記憶されていることにより表現される。なおこの点は、以降に図5(a)、図6(a)又は図7(a)を用いてそれぞれ説明するニューラルネットワークの場合においても同様である。

40

【0048】

上記ニューラルネットワークS1をモデル化する場合、図2及び図3を用いて説明した関連形態に係るニューラルネットワーク回路CSにおいて、1ビットの入力データIをn個とする。このとき、当該n個の入力データIが入力されるニューラルネットワーク回路CSのそれぞれが、図4(a)においてハッチングで示されるニューロンNRの機能をモデル化したものであり、上記乗算処理、加算処理及び活性化処理をそれぞれ実行する。な

50

お以下の図4乃至図7を用いた説明においては、上記n個の入力データIが入力されるニューラルネットワーク回路CSを、「ニューラルネットワーク回路CS1」、「ニューラルネットワーク回路CS2」、…、と称する。そして関連形態に係るニューラルネットワーク集積回路の第1例としては、当該n個の入力データIが入力されるニューラルネットワーク回路CS1等をm個集積する。

【0049】

即ち図4(b)に示すように、関連形態に係るニューラルネットワーク集積回路の第1例であるニューラルネットワーク集積回路C1は、各1ビットのn個の入力データ I_1 乃至入力データ I_n がそれぞれ共通的に入力されるm個のニューラルネットワーク回路CS1乃至ニューラルネットワーク回路CSmが集積されて構成されている。そして、ニューラルネットワーク回路CS1乃至ニューラルネットワーク回路CSmのそれぞれには、上記タイミング信号 τ_1 等がタイミング生成回路TGから共通的に入力される。このときタイミング生成回路TGは、予め設定された基準クロック信号CLKに基づいて上記タイミング信号 τ_1 等を生成してニューラルネットワーク回路CS1乃至ニューラルネットワーク回路CSmに出力する。そしてニューラルネットワーク回路CS1乃至ニューラルネットワーク回路CSmの各々は、上記入力データ I_1 乃至入力データ I_n と、タイミング信号 τ_1 等と、に基づいて、各1ビットの出力データ O_1 、出力データ O_2 、…、出力データ O_m をそれぞれ出力する。

10

【0050】

以上説明した構成を備えるニューラルネットワーク集積回路C1において、m個のニューロンNRに対してn個のニューロンNRからそれぞれ出力データOが出力されることにより、m個のニューロンNRから出力データOが合計m個出力される図4(a)のニューラルネットワークS1がモデル化される。

20

【0051】

(D) 関連形態に係るニューラルネットワーク集積回路の第2例について

次に、関連形態に係るニューラルネットワーク集積回路の第2例について、図5を用いて説明する。なお図5において、図2及び図3を用いて説明した関連形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【0052】

関連形態に係るニューラルネットワーク集積回路の第2例は、図5(a)に例示するニューラルネットワークSS1をモデル化するためのニューラルネットワーク集積回路である。当該ニューラルネットワークSS1は、図4(a)を用いて説明したニューラルネットワークS1において $n = m$ とした場合に相当する。即ちニューラルネットワークSS1は、図5(a)においてハッチングで示される $3 \times n$ 個のニューロンNRのそれぞれに対して相隣接する列の(n個の)ニューロンNRから出力データOがそれぞれ出力されることにより、図5(a)右端列のn個のニューロンNRから出力データOがそれぞれ出力されるニューラルネットワークである。上記ニューラルネットワークSS1は、並列多入力・並列多出力型の多段ニューラルネットワークである。

30

【0053】

上記ニューラルネットワークSS1をモデル化する場合も、図4を用いて説明したニューラルネットワークS1と同様に、図2及び図3を用いて説明した関連形態に係るニューラルネットワーク回路CSにおいて、1ビットの入力データIをn個とする。このとき、当該n個の入力データIが入力されるニューラルネットワーク回路CSのそれぞれが、図5(a)においてハッチングで示されるニューロンNRの機能をモデル化したものであり、上記乗算処理、加算処理及び活性化処理をそれぞれ実行する。そして関連形態に係るニューラルネットワーク集積回路の第2例としては、当該n個の入力データIが入力されるニューラルネットワーク回路CS1等を直列に接続して計 $3 \times n$ 個集積する。

40

【0054】

即ち図5(b)に示すように、関連形態に係るニューラルネットワーク集積回路の第2

50

例であるニューラルネットワーク集積回路 $CC1$ は、各 1 ビットの n 個の入力データ I_1 乃至入力データ I_n がそれぞれ共通的に入力される n 個のニューラルネットワーク回路 $CS11$ 乃至ニューラルネットワーク回路 $CS1n$ が集積されて一のニューラルネットワーク集積回路 $C1$ が構成される (図 4 (b) 参照)。そして、ニューラルネットワーク集積回路 $C1$ を構成するニューラルネットワーク回路 $CS11$ 乃至ニューラルネットワーク回路 $CS1n$ のそれぞれは、各 1 ビットの出力データ O_{11} 乃至出力データ O_{1n} を出力し、それらが次段の n 個のニューラルネットワーク回路 $CS21$ 乃至ニューラルネットワーク回路 $CS2n$ に共通的に入力される。これらニューラルネットワーク回路 $CS21$ 乃至ニューラルネットワーク回路 $CS2n$ により、他の一のニューラルネットワーク集積回路 $C2$ が構成される。そして、ニューラルネットワーク集積回路 $C2$ を構成するニューラルネットワーク回路 $CS21$ 乃至ニューラルネットワーク回路 $CS2n$ のそれぞれは、各 1 ビットの出力データ O_{21} 乃至出力データ O_{2n} を出力し、それらが次段の n 個のニューラルネットワーク回路 $CS31$ 乃至ニューラルネットワーク回路 $CS3n$ に共通的に入力される。これらニューラルネットワーク回路 $CS31$ 乃至ニューラルネットワーク回路 $CS3n$ により、更に一のニューラルネットワーク集積回路 $C3$ が構成される。ここで、各ニューラルネットワーク回路 $CS11$ 等に対しては、図 4 (a) に示す場合と同様に上記タイミング信号 ϕ_1 等が共通的に入力されているが、説明の簡略化のために図 5 (b) では図示を省略している。そしてニューラルネットワーク集積回路 $C1$ は上記入力データ I_1 乃至入力データ I_n と、タイミング信号 ϕ_1 等と、に基づいて、出力データ O_{11} 、出力データ O_{12} 、...、出力データ O_{1n} をそれぞれ生成し、これらを次段のニューラルネットワーク集積回路 $C2$ に共通的に出力する。次にニューラルネットワーク集積回路 $C2$ は上記出力データ O_{12} 乃至出力データ O_{1n} と、タイミング信号 ϕ_1 等と、に基づいて、出力データ O_{21} 、出力データ O_{22} 、...、出力データ O_{2n} をそれぞれ生成し、これらを次段のニューラルネットワーク集積回路 $C3$ に共通的に出力する。最後にニューラルネットワーク集積回路 $C3$ は上記出力データ O_{21} 乃至出力データ O_{2n} と、タイミング信号 ϕ_1 等と、に基づいて、最終的な出力データ O_{31} 、出力データ O_{32} 、...、出力データ O_{3n} をそれぞれ生成して出力する。

【0055】

以上説明した構成を備えるニューラルネットワーク集積回路 $CC1$ において、 n 個のニューロン NR から次段の n 個のニューロン NR に対してそれぞれが 1 ビットの出力データ O がそれぞれ出力されることが段階的に繰り返されることにより、最終的に出力データ O が合計 n 個出力される図 5 (a) のニューラルネットワーク $SS1$ がモデル化される。

【0056】

(E) 関連形態に係るニューラルネットワーク集積回路の第 3 例について

次に、関連形態に係るニューラルネットワーク集積回路の第 3 例について、図 6 を用いて説明する。なお図 6 において、図 2 及び図 3 を用いて説明した関連形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【0057】

関連形態に係るニューラルネットワーク集積回路の第 3 例は、図 6 (a) に例示するニューラルネットワーク $SS2$ をモデル化するためのニューラルネットワーク集積回路の例である。当該ニューラルネットワーク $SS2$ は、それぞれが図 6 (a) においてハッチングで示される m 個のニューロン NR からなる複数の組により構成され、これらのニューロン NR のそれぞれに対して共通の n 個のニューロン NR (図 6 (a) において破線で示される) から 1 ビットの出力データ O がそれぞれ出力されることにより、図 6 (a) においてハッチングで示される各ニューロン NR から各 1 ビットで合計 $m \times$ 組数個の出力データ O が出力されるニューラルネットワークである。このニューラルネットワーク $SS2$ の場合、図 6 (a) においてハッチングで示される各ニューロン NR は、各 1 ビットで同じ数 (n 個) の出力データ O をそれぞれ受信することになる。即ち上記ニューラルネットワーク $SS2$ は、並列多入力 - 並列多出力型の一段ニューラルネットワークである。

【 0 0 5 8 】

上記ニューラルネットワーク S S 2 をモデル化する場合も、図 4 を用いて説明したニューラルネットワーク S 1 と同様に、図 2 及び図 3 を用いて説明した関連形態に係るニューラルネットワーク回路 C S において、1 ビットの入力データ I を n 個とする。このとき、当該 n 個の入力データ I が入力されるニューラルネットワーク回路 C S のそれぞれが、図 6 (a) においてハッチングで示されるニューロン N R の機能をモデル化したものであり、上記乗算処理、加算処理及び活性化処理をそれぞれ実行する。そして関連形態に係るニューラルネットワーク集積回路の第 3 例としては、当該 n 個の入力データ I が入力されるニューラルネットワーク回路 C S 1 1 等を並列に接続して上記組数分集積する。

【 0 0 5 9 】

即ち図 6 (b) に示すように、関連形態に係るニューラルネットワーク集積回路の第 3 例であるニューラルネットワーク集積回路 C C 2 は、各 1 ビットの n 個の入力データ I_1 乃至入力データ I_n がそれぞれ共通的に入力される m 個のニューラルネットワーク回路 C S 1 1 乃至ニューラルネットワーク回路 C S 1 m が集積されて、一のニューラルネットワーク集積回路 C 1 が構成される (図 4 (b) 参照)。また同じ n 個の入力データ I_1 乃至入力データ I_n がそれぞれ並列的且つ共通的に入力される m 個のニューラルネットワーク回路 C S 2 1 乃至ニューラルネットワーク回路 C S 2 m が集積されて、他の一のニューラルネットワーク集積回路 C 2 が構成される (図 4 (b) 参照)。これ以降、同様に n 個の入力データ I_1 乃至入力データ I_n がそれぞれ並列的且つ共通的に入力される m 個のニューラルネットワーク回路が集積されて、図 6 (b) において図示を省略する他のニューラルネットワーク集積回路がそれぞれ構成される。ここで、各ニューラルネットワーク回路 C S 1 1 等に対しては、図 5 を用いて説明した場合と同様に、図 4 (a) に示す場合と同様の上記タイミング信号 τ_1 等が共通的に入力されているが、説明の簡略化のために図 6 (b) では図示を省略している。そしてニューラルネットワーク集積回路 C 1 は上記入力データ I_1 乃至入力データ I_n と、タイミング信号 τ_1 等と、に基づいて、各 1 ビットの出力データ $O_{1,1}$ 、出力データ $O_{1,2}$ 、...、出力データ $O_{1,m}$ をそれぞれ生成して出力する。一方ニューラルネットワーク集積回路 C 2 は、同じ入力データ I_1 乃至入力データ I_n と、タイミング信号 τ_1 等と、に基づいて、各 1 ビットの出力データ $O_{2,1}$ 、出力データ $O_{2,2}$ 、...、出力データ $O_{2,m}$ をそれぞれ生成して出力する。これ以降、図示を省略する他のニューラルネットワーク集積回路も、それぞれに m 個の出力データを出力する。

【 0 0 6 0 】

以上説明した構成を備えるニューラルネットワーク集積回路 C C 2 において、m x 組数分のニューロン N R から並行的にそれぞれ出力データ O が出力されることにより、最終的に出力データ O が合計 m x 組数分出力される図 6 (a) のニューラルネットワーク S S 2 がモデル化される。

【 0 0 6 1 】

(F) 関連形態に係るニューラルネットワーク集積回路の第 4 例について

最後に、関連形態に係るニューラルネットワーク集積回路の第 4 例について、図 7 を用いて説明する。なお図 7 において、図 2 及び図 3 を用いて説明した関連形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【 0 0 6 2 】

関連形態に係るニューラルネットワーク集積回路の第 4 例は、図 7 (a) に例示するニューラルネットワーク S S 3 をモデル化するためのニューラルネットワーク集積回路の例である。当該ニューラルネットワーク S S 3 は、これまで説明してきた関連形態に係るニューラルネットワーク S 1 等に対して、ニューロン N R の数及びニューロン N R 間の接続の態様についての自由度を更に向上させたニューラルネットワークである。なお図 7 (a) には、段階的に各 1 ビットの出力データ O (入力データ I) が授受されるニューロン群 (図 7 (a) 破線参照) に属するニューロン N R の数が相互に異なるニューラルネットワーク S S 3 について例示している。

10

20

30

40

50

【 0 0 6 3 】

上記ニューラルネットワーク S S 3 をモデル化する場合、図 2 乃至図 3 を用いて説明した関連形態に係るニューラルネットワーク回路 C S において、1 ビットの入力データ I を例えば n 個とする。このとき、当該 n 個の入力データ I が入力されるニューラルネットワーク回路 C S のそれぞれが、図 7 (a) に示す各ニューロン N R の機能をモデル化したものであり、上記乗算処理、加算処理及び活性化処理をそれぞれ実行する。そして関連形態に係るニューラルネットワーク集積回路の第 4 例としては、当該 n 個の入力データ I が入力されるニューラルネットワーク回路 C S 1 1 等をそれぞれに複数備えたニューラルネットワーク集積回路を複数備え、当該各ニューラルネットワーク集積回路を、後述する複数のスイッチ及びそれらを切り換えるスイッチボックスにより接続して集積する。

10

【 0 0 6 4 】

即ち図 7 (b) に示すように、関連形態に係るニューラルネットワーク集積回路の第 4 例であるニューラルネットワーク集積回路 C C 3 は、各 1 ビットの n 個の入力データ I_1 乃至入力データ I_n がそれぞれ共通的に入力される n 個のニューラルネットワーク回路 C S 1 1 乃至ニューラルネットワーク回路 C S 1 n が集積されて一のニューラルネットワーク集積回路 C 1 が構成される (図 4 (b) 参照)。そして同様に、例えば m 個のニューラルネットワーク回路 C S 2 1 乃至ニューラルネットワーク回路 C S 2 m が集積されて一のニューラルネットワーク集積回路 C 2 が、ニューラルネットワーク回路 C S 3 1 乃至ニューラルネットワーク回路 C S 3 p (p は 2 以上の自然数。以下、同様。) が集積されて一のニューラルネットワーク集積回路 C 3 が、ニューラルネットワーク回路 C S 4 1 乃至ニューラルネットワーク回路 C S 4 q (q は 2 以上の自然数。以下、同様。) が集積されて一のニューラルネットワーク集積回路 C 4 が、それぞれ構成される。また、ニューラルネットワーク集積回路 C 1 乃至ニューラルネットワーク集積回路 C 4 のそれぞれは、図 7 (b) に例示するようにスイッチ S W 1 乃至スイッチ S W 4 を介して、相互に各 1 ビットの入力データ I 及び出力データ O の授受が可能とされている。そして、ニューラルネットワーク集積回路 C 1 乃至ニューラルネットワーク集積回路 C 4 間における入力データ I 及び出力データ O の授受の態様 (即ちニューラルネットワーク集積回路 C 1 乃至ニューラルネットワーク集積回路 C 4 間の接続態様) が、スイッチ S W 1 乃至スイッチ S W 4 を介してスイッチボックス S B 1 乃至スイッチボックス S B 4 により切り換えられる。

20

【 0 0 6 5 】

次に、上記スイッチボックス S B 1 乃至スイッチボックス S B 4 の細部構成について、図 7 (c) を用いて説明する。なお、スイッチボックス S B 1 乃至スイッチボックス S B 4 はいずれも同様の構成を備えるので、図 7 (c) では、これらを纏めてスイッチボックス S B として説明する。

30

【 0 0 6 6 】

図 7 (c) に示すように、ニューラルネットワーク集積回路 C C 3 における各 1 ビットの入力データ I 又は出力データ O の接続態様、及び結果的に有効なニューロン N R の数を制御するスイッチボックス S B は、セレクトタ M_1 乃至セレクトタ M_5 が図 7 (c) に示す態様により接続されて構成されている。この図 7 (c) に示すスイッチボックス S B の構成においては、上述してきた入力データ I に対応する信号は図 7 (c) 中左から入力される信号であり、上記出力データ O に対応する信号は図 7 (c) 中の上方及び下方からそれぞれ入力される信号である。そしてニューラルネットワーク集積回路 C 1 乃至ニューラルネットワーク集積回路 C 4 に対する入力データ I 等の切り換えは、当該切り換えを制御する切替制御信号 S_{c1} 乃至切替制御信号 S_{c5} が外部からそれぞれ入力されるセレクトタ M_1 乃至セレクトタ M_5 により、実行される。

40

【 0 0 6 7 】

以上説明したように、入力データ I に対応する出力データ O を生成して出力する図 7 (a) のニューラルネットワーク S S 3 が、図 7 (c) に示す構成を備えるスイッチボックス S B 1 乃至スイッチボックス S B 4 により各スイッチ S W 1 乃至スイッチ S W 4 が切り換えられる、図 7 (b) に示す構成のニューラルネットワーク集積回路 C C 3 によりモデ

50

ル化される。

【0068】

以上それぞれ説明したように、関連形態に係るニューラルネットワーク回路CS及びニューラルネットワーク集積回路C1等の構成及び動作によれば、図2及び図3に例示するように、対応すべき脳機能に基づいてその数が既定されているメモリセル1のそれぞれが、記憶値として「NC」を意味する所定値或いは「1」又は「0」のいずれかを記憶し、1ビットの入力データIの値と記憶値とが等しい場合に入力データIの入力に対応して「1」を出力し、入力データIの値と記憶値とが等しくない場合に入力データIの入力に対応して「0」を出力し、上記所定値が記憶されている場合にいずれの値の入力データIが入力されても当該所定値を出力する。そして多数判定回路2が、値「1」を出力するメモリセル1の総数が値「0」を出力するメモリセル1の総数より大きい場合に値「1」を出力データOとして出力し、値「1」を出力するメモリセル1の総数が値「0」を出力するメモリセル1の総数以下の場合に「0」を出力データOとして出力する。よって、メモリセル1においてニューラルネットワーク回路としての乗算処理を行い、一の多数判定回路2によりニューラルネットワーク回路としての加算処理及び活性化処理を行うので、回路規模及びそれに対応するコストを大幅に縮小しつつ、ニューラルネットワーク回路を効率的に実現することができる。

10

【0069】

また図4(b)に例示するように、それぞれが1ビットのn個の入力データIにそれぞれ対応してメモリセル1の数がnであるニューラルネットワーク回路CSをm個備え、各ニューラルネットワーク回路CSに対してn個の入力データIが並列且つ共通的に入力され、各ニューラルネットワーク回路CSから出力データOをそれぞれ出力させる場合は、図4(a)に例示するニューラルネットワークS1をモデル化し且つ入力がn個であり出力がm個であるn×mのニューラルネットワーク集積回路C1を、回路規模及びそれに対応するコストを大幅に縮小しつつ効率的に実現することができる。更にこの場合に、図4(a)においてハッチングで示されるm個のニューロンNRと、当該m個に対して出力データOをそれぞれ出力するn個のニューロンNRと、の間に多様な接続のパターンがあっても、ニューラルネットワーク集積回路C1においてニューロンNR間に接続がない場合に対応するメモリセル1の記憶値として上記所定値を用いることで、ニューラルネットワーク集積回路C1をより効率的に実現することができる。更に図4に例示する場合には、n個の入力データIを各ニューラルネットワーク回路CSに対して並列且つ共通的に入力し、それに基づくm個の出力データOを並列に出力させることができるため、入力データI及び出力データOを逐次的に入出力させなければならない場合と比べて処理の大幅な高速化が可能となる。

20

30

【0070】

更にまた図5に例示するように、上記「n」と上記「m」とが等しいニューラルネットワーク集積回路C1等を直列に接続し、一のニューラルネットワーク集積回路C1(又はニューラルネットワーク集積回路C2)からの出力データOを、当該ニューラルネットワーク集積回路C1(又はニューラルネットワーク集積回路C2)の直後に接続され他のニューラルネットワーク集積回路C2(又はニューラルネットワーク集積回路C3)における入力データIとする場合は、入力及び出力が共に並列型であるニューラルネットワーク集積回路C1を、回路規模及びそれに対応するコストを大幅に縮小しつつ効率的に実現することができる。

40

【0071】

また図6に例示するように、各ニューラルネットワーク集積回路CSに対してn個の入力データIが並列且つ共通に入力され、各ニューラルネットワーク集積回路CSからm個の出力データOを並列にそれぞれ出力する場合は、入力及び出力が共に並列型であり且つ出力データOの数が入力データIの数より多いニューラルネットワーク集積回路C2を、回路規模及びそれに対応するコストを大幅に縮小しつつ効率的に実現することができる。

50

【 0 0 7 2 】

また図7に例示するように、ニューラルネットワーク集積回路C1等を複数備え、各ニューラルネットワーク集積回路C1等をアレイ状に且つ相互に接続するスイッチSW1等により、各ニューラルネットワーク集積回路C1等に対する入力データI及び出力データOが切り換える場合には、スイッチSW1等における切換動作が、対応すべき脳機能に基づいて既定されていれば、大規模なニューラルネットワーク集積回路CC3を、対応するコストを大幅に縮小しつつ効率的に実現することができる。

【 0 0 7 3 】

(II) 実施形態

次に、本発明に係る実施形態について、図8乃至図16を用いて説明する。なお、図8及び図9は実施形態に係るニューラルネットワーク集積回路の第1例をそれぞれ示す図であり、図10は実施形態に係るニューラルネットワーク回路の第1例を示す図であり、図11は実施形態に係るニューラルネットワーク集積回路の第2例を示す図である。また、図12は当該ニューラルネットワーク集積回路の第3例を示す図であり、図13は当該ニューラルネットワーク集積回路の第4例を示す図であり、図14は当該第4例の細部構成を示す図であり、図15は実施形態に係るニューラルネットワーク集積回路の第5例を示す図であり、図16は当該ニューラルネットワーク集積回路の第6例等の構成を示すブロック図である。

10

【 0 0 7 4 】

以下に説明する実施形態は、図1乃至図7を用いて上述してきたニューラルネットワークS等のモデル化の構成又は手法とは異なる構成又は手法により、当該ニューラルネットワークS等をモデル化しようとするものである。

20

【 0 0 7 5 】

(A) 実施形態に係るニューラルネットワーク集積回路の第1例について

初めに、実施形態に係るニューラルネットワーク集積回路の第1例について、図8及び図9を用いて説明する。なお、図8は当該第1例としての上記乗算処理を行う当該第1例の一部を示す図であり、図9は当該第1例全体を示す図である。このとき、当該乗算処理が本発明に係る「乗算機能」の一例に相当する。

【 0 0 7 6 】

図8(a)に例示するように、当該第1例の一部によりモデル化されるネットワークS'では、一つのニューロンNRから1ビットの出力データO(換言すれば入力データI)が入力される。そして、入力データIの出力先となる図示しない複数の他のニューロンにそれぞれ対応する異なった重み付け係数 W_1 乃至重み付け係数 W_4 のうちの一つが当該入力データIに乘算され、上記図示しない他のニューロンに対して、出力データ E_1 乃至出力データ E_4 としてそれぞれ出力される。そしてこのときの出力データEは、入力データIと同様に1ビットの信号である。よって、図8に示す入力データIの値、各重み付け係数Wの値及び出力データEの値は、いずれも「0」又は「1」のいずれかである。

30

【 0 0 7 7 】

次に、実施形態に係るニューラルネットワーク集積回路の第1例における図8(a)に示すネットワークS'に相当する部分の構成を、図8(b)にネットワーク回路CS'として示す。当該ネットワーク回路CS'は、各々が図8(a)に示す出力データ E_1 乃至出力データ E_4 にそれぞれ対応する四組のメモリセル10及びメモリセル11と、各々が出力データE(換言すれば、上記図示しない他のニューロンの入力データI)にそれぞれ対応する四つの多数判定入力回路12と、により構成される。このとき、一つのメモリセル10と一つのメモリセル11からなるメモリセル対の数及びそれらに対応する多数判定入力回路12の数(図8に例示する場合は共に四つ)は、実施形態に係るニューラルネットワーク集積回路の第1例として所望される出力データOの数に等しい。なお以下の図8の説明において、出力データOの数分の上記メモリセル対を、纏めて「メモリセルブロック15」(図8(b)破線参照)と称する。また、メモリセル10が本発明に係る「第1記憶部」の一例に相当し、メモリセル11が本発明に係る第2記憶部の一例に相当し、多

40

50

数判定入力回路 1 2 が本発明に係る「第 1 演算部」の一例に相当する。

【 0 0 7 8 】

以上の構成において各メモリセルブロック 1 5 内のメモリセル 1 0 は、ネットワーク回路 C S ' を含む実施形態に係るニューラルネットワーク集積回路の第 1 例が対応すべき脳機能に基づいて予め設定された 1 ビットの重み付け係数 W をそれぞれに記憶している。これに対して各メモリセルブロック 1 5 内のメモリセル 1 1 は、上記脳機能に基づいて予め設定された 1 ビットの接続有無情報をそれぞれに記憶している。ここで当該接続有無情報は、上記関連形態におけるメモリセル 1 の記憶値「NC」に相当するものであり、実施形態に係るニューラルネットワークにおける二つのニューロン NR 間に接続がある状態であるか、又は当該接続がない状態であるか、のいずれかを表すための記憶値である。なお、各メモリセル 1 0 及びメモリセル 1 1 にどのような記憶値を記憶させておくかは、例えば、ネットワーク S ' を含む実施形態に係るニューラルネットワーク集積回路の第 1 例としてどのような脳機能をモデル化するか等に基づいて予め設定されていけばよい。

10

【 0 0 7 9 】

そして各メモリセル 1 0 は、それぞれの記憶値を、重み付け係数 W₁、重み付け係数 W₂、重み付け係数 W₃ 及び重み付け係数 W₄ として多数判定入力回路 1 2 にそれぞれ出力する。このとき各メモリセル 1 0 は、それぞれの記憶値を、重み付け係数 W₁ 乃至重み付け係数 W₄ として同時に多数判定入力回路 1 2 に出力する。なおこの同時出力の構成は、以下の図 9 乃至図 1 6 を用いてそれぞれ説明するニューラルネットワーク回路及びニューラルネットワーク集積回路における各メモリセル 1 0 においても同様である。一方各メモリセル 1 1 も、それぞれの記憶値を、接続有無情報 C₁、接続有無情報 C₂、接続有無情報 C₃ 及び接続有無情報 C₄ として多数判定入力回路 1 2 にそれぞれ出力する。このとき各メモリセル 1 1 は、それぞれの記憶値を、接続有無情報 C₁ 乃至接続有無情報 C₄ として同時に多数判定入力回路 1 2 に出力する。また各メモリセル 1 1 は、上記各メモリセル 1 0 からの記憶値の出力に対して例えば一サイクル前又は後にずらして、それぞれの記憶値を同時に多数判定入力回路 1 2 に出力する。なおこの同時出力の構成及び各メモリセル 1 0 からの記憶値の出力のタイミングとの関係は、以下の図 9 乃至図 1 4 を用いてそれぞれ説明するニューラルネットワーク回路及びニューラルネットワーク集積回路における各メモリセル 1 1 においても同様である。更に以下の説明において、接続有無情報 C₁、接続有無情報 C₂、接続有無情報 C₃、...、に共通の事項を説明する場合、単に「接続有無情報 C」と称する。

20

30

【 0 0 8 0 】

他方各多数判定入力回路 1 2 には、図 8 (b) において図示しない他のノード NR (図 8 (a) 参照) からの 1 ビットの入力データ I が共通的に入力されている。そして各多数判定入力回路 1 2 は、対応するメモリセル 1 1 から出力される上記接続有無情報を、そのまま接続有無情報 C₁ 乃至接続有無情報 C₄ としてそれぞれ出力する。

【 0 0 8 1 】

これらに加えて各多数判定入力回路 1 2 は、対応するメモリセル 1 0 から出力される重み付け係数 W₁、重み付け係数 W₂、重み付け係数 W₃ 及び重み付け係数 W₄ と上記入力データ I との排他的否定論理和 (X N O R) を算出し、上記出力データ E₁、出力データ E₂、出力データ E₃ 及び出力データ E₄ としてそれぞれ出力する。このとき、対応するメモリセル 1 1 の記憶値 (重み付け係数 W) と、入力データ I の値と、多数判定入力回路 1 2 から出力される出力データ E の値と、の関係は、図 8 (c) に真理値表を示す関係となる。なお図 8 (c) には、上記排他的否定論理和 (X N O R) を算出する前提としての排他的論理和 (X O R) についても記載している。

40

【 0 0 8 2 】

ここで、図 2 を用いて説明した関連形態に係るニューラルネットワーク回路 C S に対応する真理値表 (図 2 (c) 参照) と、上記図 8 (c) に示す真理値表とを比較する。このとき、メモリセル 1 0 における記憶値及び入力データ I の値をそれぞれ図 2 (c) に示した真理値表のものと同じとした場合、図 8 (b) に示す出力データ E の値は図 2 (b) に

50

示す出力データEの値と同一となる。これらにより、図8(b)に示すネットワーク回路CS'は、図2(b)に示したニューラルネットワーク回路CSにおける上記乗算処理と同様の論理により、図8(a)に示すネットワークS'における上記乗算処理をモデル化した回路となる。即ち、各メモリセル10から出力されてくる各々の記憶値(重み付け係数W)と入力データIの値との間の排他的論理和を多数判定入力回路12において算出することが上記乗算処理に相当する。以上の通り、図8(b)に示すネットワーク回路CS'により、図8(a)に示すネットワークS'における乗算処理がモデル化される。

【0083】

次に、実施形態に係るニューラルネットワーク集積回路の第1例の全体について図9を用いて説明する。なお図9において、図8を用いて説明した実施形態に係るネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

10

【0084】

図9を用いてその全体を説明する実施形態に係るニューラルネットワーク集積回路の第1例は、図8を用いて説明した実施形態に係るネットワーク回路CS'を複数集積した集積回路である。実施形態に係るニューラルネットワーク集積回路の第1例では、当該ネットワーク回路CS'に相当する上記乗算処理に加えて、上記加算処理及び上記活性化処理が実行される。このとき、当該加算処理及び活性化処理が本発明に係る「加算/適用機能」の一例に相当する。

【0085】

先ず、実施形態に係るニューラルネットワーク集積回路の第1例によりモデル化されるニューラルネットワークの全体について、図9(a)を用いて説明する。図9(a)に示す当該ニューラルネットワークS1'は、図8を用いて説明したネットワークS'をm個のニューロンNR分含んでいる。当該ニューラルネットワークS1'では、図9(a)においてハッチングで示されるn個のニューロンNRのそれぞれに対して、それぞれが上記ネットワークS'を構成するm個のニューロンNRからそれぞれ1ビットの出力データO(換言すれば入力データI)が出力される。そしてこれにより、各出力データOが出力データEとなって上記ハッチングで示されるn個のニューロンNRのそれぞれに入力され、当該ハッチングで示されるニューロンNRから出力データOが一つずつ合計n個並列的に出力される。即ち上記ニューラルネットワークS1'は、直列(m)入力-並列(n)出力型の一段ニューラルネットワークである。

20

30

【0086】

上記ニューラルネットワークS1'をモデル化した実施形態に係るニューラルネットワーク集積回路の第1例は、図9(b)に示すニューラルネットワーク集積回路C1'となる。当該ニューラルネットワーク集積回路C1'は、上記メモリセル対及び多数判定入力回路12をそれぞれにn個ずつ含む実施形態に係るニューラルネットワーク回路CS'(図8参照)をm個備えると共に、各多数判定入力回路12及びメモリセル対に対応させてn個の直列多数判定回路13を備えている。このとき、直列多数判定回路13が本発明に係る「第2演算部」の一例に相当する。そして図9(b)に示すように、 $n \times m$ 個の上記メモリセル対(換言すれば、m個のメモリセルブロック15)により、メモリセルアレイMC1が構成されている。またニューラルネットワーク集積回路C1'では、図9(b)に示すメモリセルアレイMC1における横一行(m個)のメモリセル対で一つの多数判定入力回路12を共用する。なお、メモリセルアレイMC1、各多数判定入力回路12及び各直列多数判定回路13には、それぞれ上記タイミング信号 ϕ_1 等が共通的に入力されているが、説明の簡略化のために図9(b)では図示を省略している。

40

【0087】

以上の構成において、それぞれのニューラルネットワーク回路CS'を構成するメモリセルブロック15のメモリセル10からは、上記重み付け係数Wが、一のメモリセルブロック15に含まれる各メモリセル10について同時に、且つm個のメモリセルブロック15について順次に(即ちシリアル形式で)、それぞれ出力される。そして、これらの重み付け係数Wと、対応するタイミングによりシリアル形式で入力されるm個の入力データI

50

(各1ビットの入力データI)と、の上記排他的論理和を、上記共有する多数判定入力回路12において時分割的に演算し、それを出力データEとして、対応する直列多数判定回路13にシリアル形式で出力する。一方、それぞれのニューラルネットワーク回路CS'を構成するメモリセルブロック15のメモリセル11からは、上記接続有無情報Cが、一のメモリセルブロック15に含まれる各メモリセル11について同時に、且つm個のメモリセルブロック15について順次に(即ちシリアル形式で)、それぞれ出力される。そしてこれらの接続有無情報Cが、上記共有する多数判定入力回路12を介し、各入力データIの入力タイミングに対応したシリアル形式で、対応する直列多数判定回路13に出力される。なお、各メモリセルブロック15からの上記各重み付け係数Wの出力タイミングの態様、及び各メモリセルブロック15からの上記各接続有無情報Cの出力タイミングの態様のそれぞれは、以下の図10乃至図14を用いてそれぞれ説明するニューラルネットワーク集積回路における各メモリセル11においても同様である。また図9(b)では、図面としての明確化及びその説明の簡略化のため、メモリセル10から出力される上記重み付け係数 W_n と、それに対応するメモリセル11から出力される上記接続有無情報 C_n と、を、合わせて一本の実線で示している。この点は、以下に説明する図11乃至図14において同様である。

10

【0088】

次に、各多数判定入力回路12から出力データE及び接続有無情報Cがそれぞれ入力されるn個の直列多数判定回路13はそれぞれ、同じタイミングで入力された接続有無情報Cが「接続あり」を示している最大m個の出力データEについて、値「1」の当該出力データEの数を加算してその合計値を算出すると共に、値「0」の出力データEの数を加算してその合計値を算出する。これらの加算が上記加算処理に相当する。そして各直列多数判定回路13はそれぞれ、値「1」の出力データE及び値「0」の出力データEそれぞれの数の上記合計値を比較し、前者の数から後者の数を減じた値が、関連形態に係る上記多数判定閾値と同様にして予め設定された多数判定閾値以上の場合にのみ、値「1」の出力データOを出力する。一方それ以外の場合、即ち値「1」の出力データEの数の合計値から値「0」の出力データEの数の合計値を減じた値が上記多数判定閾値未満の場合に各直列多数判定回路13はそれぞれ、値「0」の出力データOを出力する。これらの各直列多数判定回路13における処理が上記活性化処理に相当すると共に、各出力データOは1ビットとなる。ここで、同じタイミングで出力される上記接続有無情報Cが「接続なし」を示している場合、直列多数判定回路13は、出力データEを、値「1」の出力データEの数及び値「0」の出力データEの数のいずれにも加算しない。そして各直列多数判定回路13は、上述した各処理により1ビットの出力データOを出力することを、入力データIが入力されるタイミングに合わせて繰り返す。このときの出力データOは、結果的に各直列多数判定回路13から並列的に出力される。この場合、出力データOの総数はn個となる。以上の通り、図9(a)においてハッチングで示される一つのニューロンNRに対応する上記乗算処理、加算処理及び活性化処理のそれぞれは、図9(b)に示すメモリセルアレイMC1における横一行分のメモリセル対と、それらに対応する多数判定入力回路12及び直列多数判定回路13と、により実行されることになる。

20

30

【0089】

以上説明したように、図9(a)においてハッチングで示されるn個のニューロンNRに対してm個のニューロンNRから1ビットの出力データOがそれぞれ出力され、これらにより当該n個のニューロンNRから出力データOが合計n個出力されるニューラルネットワークS1'が、図9(b)に示す構成を備えるニューラルネットワーク集積回路C1'によりモデル化される。

40

【0090】

より具体的に、人の脳機能に対応し且つ各1ビットの入力データI及び出力データOがそれぞれ入出力されるニューラルネットワーク集積回路C1'の構成によれば、上記乗算処理をメモリセル10、メモリセル11及び多数判定入力回路12により実行し、上記加算処理及び活性化処理を直列多数判定回路13により実行するので、回路規模及びそれに

50

対応するコストを大幅に縮小しつつ、人の脳機能に対応したニューラルネットワーク S 1 ' を効率的に実現することができる。

【 0 0 9 1 】

また、上記ニューラルネットワーク集積回路 C 1 ' の構成によれば、入力データ I の数が一であり、出力データ O、メモリセル 1 0、メモリセル 1 1、多数判定入力回路 1 2 及び直列多数判定回路 1 3 それぞれの数が、対応すべき脳機能に基づいた既定の二以上の数とされており、重み付け係数 W 及び接続有無情報 C のそれぞれが脳機能に対応した値とされている。そして各多数判定入力回路 1 2 が、各出力データ O にそれぞれ対応する重み付け係数 W と入力データ I との排他的論理和を出力データ O ごとにそれぞれ生成し、各直列多数判定回路 1 3 が、(値「 1 」である出力データ E の総数 - 値「 0 」である出力データ E の総数 > 多数判定閾値) であるとき値「 1 」を出力データ O として出力し、(値「 1 」である出力データ E の総数 - 値「 0 」である出力データ E の総数 < 多数判定閾値) であるとき値「 0 」を出力データ O として出力する。よって、各メモリセル 1 0、各メモリセル 1 1 及び各多数判定入力回路 1 2 により出力データ O ごとに乗算処理を実行し、更に各直列多数判定回路 1 3 により出力データ O ごとに加算処理及び活性化処理を実行するので、回路規模及びそれに対応するコストを大幅に縮小しつつ、いわゆる直列入力 - 並列出力型のニューラルネットワーク S 1 ' を効率的に実現することができる。

10

【 0 0 9 2 】

(B) 実施形態に係るニューラルネットワーク回路の第 1 例について

次に、実施形態に係るニューラルネットワーク回路の第 1 例について、図 1 0 を用いて説明する。

20

【 0 0 9 3 】

図 1 0 (a) に例示するように、当該第 1 例に相当するニューラルネットワーク S は、基本的には図 2 (a) に例示した関連形態に係るニューラルネットワーク S と同一の構成である。但し図 1 0 (a) に示す例では、図 1 0 (a) においてハッチングで示される一つのニューロン N R に対して他の三つのニューロン N R から 1 ビットの入力データ I (当該他のニューロン N R から見ると出力データ O) が並列的に入力され、それに対応する出力データ O が当該ニューロン N R からシリアル形式で一つ出力される構成となっている。このときの出力データ O も、各入力データ I と同様に 1 ビットの信号である。よって、図 1 0 に示す入力データ I の値及び出力データ O の値は、いずれも「 0 」又は「 1 」のいずれかである。そして、図 1 0 (a) に示すニューロン N R において実行される上記乗算処理等に相当する上記式 (1) は、上記式 (1) において $n = 3$ とした場合の式である。即ち上記ニューラルネットワーク S は、並列入力 - 直列出力型の一段ニューラルネットワークである。

30

【 0 0 9 4 】

次に、図 1 0 (a) においてハッチングで示したニューロン N R に相当する実施形態に係るニューラルネットワーク回路の第 1 例の構成を、図 1 0 (b) にニューラルネットワーク回路 C C S ' として示す。当該ニューロン N R に相当する実施形態に係るニューラルネットワーク回路 C C S ' は、各々が図 1 0 (a) に例示する入力データ I にそれぞれ対応する三組のメモリセル 1 0 及びメモリセル 1 1 と、各入力データ I が入力される並列多数判定回路 2 0 と、により構成される。このとき、一つのメモリセル 1 0 と一つのメモリセル 1 1 からなるメモリセル対の数及びそれらに対応する多数判定入力回路 1 2 の数 (図 1 0 に例示する場合は共に三つ) は、図 1 0 (a) に示すニューラルネットワーク S として所望される入力データ I の数に等しい。なお以下の図 1 0 の説明において、入力データ I の数分の上記メモリセル対はそれぞれメモリセルブロック 1 5 と示されている (図 1 0 (b) 破線参照) 。

40

【 0 0 9 5 】

以上の構成において各メモリセルブロック 1 5 内のメモリセル 1 0 は、ニューラルネットワーク回路 C C S ' が対応すべき脳機能に基づいて予め設定された 1 ビットの重み付け係数 W をそれぞれに記憶している。これに対して各メモリセルブロック 1 5 内のメモリセ

50

ル 11 は、上記脳機能に基づいて予め設定された 1 ビットの接続有無情報をそれぞれに記憶している。ここで当該接続有無情報は、図 8 及び図 9 を用いて説明した実施形態に係るニューラルネットワーク回路の第 1 例における接続有無情報 C_n と同一の情報であるので、細部の説明は省略する。また、各メモリセル 10 及びメモリセル 11 にどのような記憶値を記憶させておくかは、例えば、図 10 (a) に示すニューラルネットワーク S としてどのような脳機能をモデル化するかに基づいて予め設定されていけばよい。

【0096】

そして各メモリセル 10 は、それぞれの記憶値を、図 8 (b) に示す各メモリセル 10 と同様のタイミングで、重み付け係数 W_1 、重み付け係数 W_2 及び重み付け係数 W_3 として並列多数判定回路 20 にそれぞれ出力する。一方各メモリセル 11 も、それぞれの記憶値である接続有無情報 C を、図 8 (b) に示す各メモリセル 11 と同様のタイミングで並列多数判定回路 20 にそれぞれ出力する。

10

【0097】

他方並列多数判定回路 20 には、上述したように入力データ I_1 、入力データ I_2 及び入力データ I_3 (各 1 ビット) が並列的に入力されている。そして並列多数判定回路 20 は、図 9 を用いて説明した一組の多数判定入力回路 12 及び直列多数判定回路 13 と同様の動作を含む動作 (即ち、上記乗算処理、上記加算処理及び上記活性化処理) を行う。具体的に並列多数判定回路 20 は先ず、対応する接続有無情報 C が「接続あり」を示している場合に、それぞれが 1 ビットの各入力データ I と、それらに対応する重み付け係数 W と、の上記排他的論理和を当該各入力データ I について演算する。次に並列多数判定回路 20 は、各上記演算結果について、値「1」の当該演算結果の数を加算してその合計値を算出すると共に、値「0」の当該演算結果の数を加算してその合計値を算出する。そして並列多数判定回路 20 は、値「1」の当該演算結果及び値「0」の当該演算結果それぞれの数の上記合計値を比較し、前者の数から後者の数を減じた値が、関連形態に係る上記多数判定閾値と同様にして予め設定された多数判定閾値以上の場合にのみ、値「1」の出力データ O をシリアル形式で出力する。一方それ以外の場合、即ち値「1」の出力データ E の数の合計値から値「0」の出力データ E の数の合計値を減じた値が上記多数判定閾値未満の場合に並列多数判定回路 20 は値「0」の出力データ O をシリアル形式で出力する。この場合に出力データ O は 1 ビットとなる。ここで、対応する接続有無情報 C が「接続なし」を示している場合、並列多数判定回路 20 は上記排他的論理和を演算しない。なお、各入力データ I と、対応する重み付け係数 W と、の上記排他的論理和を全ての入力データ I について一旦演算し、対応する接続有無情報 C が「接続なし」を示している場合に、その演算結果を値「1」の演算結果の数及び値「0」の演算結果の数のいずれにも加算しないように構成してもよい。そして並列多数判定回路 20 は、上述した各処理により 1 ビットの出力データ O をシリアル形式で出力することを、並列的に入力されている各入力データ I の数ごとに繰り返す。以上の各処理により、図 10 (b) に示すニューラルネットワーク回路 CCS' は、図 10 (a) においてハッチングで示されるニューロン NR における上記乗算処理、加算処理及び活性化処理をモデル化した回路となる。

20

30

【0098】

(C) 実施形態に係るニューラルネットワーク集積回路の第 2 例について

40

次に、実施形態に係るニューラルネットワーク集積回路の第 2 例について、図 11 を用いて説明する。なお図 11 において、図 10 を用いて説明した実施形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【0099】

図 11 を用いて説明する実施形態に係るニューラルネットワーク集積回路の第 2 例は、図 10 を用いて説明した実施形態に係るニューラルネットワーク回路 CCS' を複数集積した集積回路であり、より多くのニューロン NR からなる複雑なニューラルネットワークをモデル化するためのものである。

【0100】

50

先ず、実施形態に係るニューラルネットワーク集積回路の第2例によりモデル化されるニューラルネットワークについて、図11(a)を用いて説明する。図11(a)に示す当該ニューラルネットワークS2'は、図11(a)においてハッチングで示されるm個のニューロンNRのそれぞれに対してn個のニューロンNRから1ビットの出力データO(m個のニューロンNRから見た場合は入力データI)が並列的にそれぞれ入力され、それらに対応する出力データOが当該ニューロンNRからシリアル形式で出力される構成となっている。このときの出力データOも、各入力データIと同様に1ビットの信号である。よって、図11に示す入力データIの値及び出力データOの値は、いずれも「0」又は「1」のいずれかである。即ち上記ニューラルネットワークS2'は、並列入力・直列出力型の一段ニューラルネットワークである。

10

【0101】

上記ニューラルネットワークS2'をモデル化した実施形態に係るニューラルネットワーク集積回路の第2例は、図11(b)に示すニューラルネットワーク集積回路C2'となる。当該ニューラルネットワーク集積回路C2'は、上記メモリセル対をそれぞれにn個ずつ含む実施形態に係るニューラルネットワーク回路CCS'(図10参照)をm個備えると共に、上記並列多数判定回路20を備えている。そして図11(b)に示すように、n×m個の上記メモリセル対(換言すれば、m個のメモリセルブロック15)により、メモリセルアレイMC2が構成されている。またニューラルネットワーク集積回路C2'では、一つの並列多数判定回路20を図11(b)に示すメモリセルアレイMC2における横一行(m個)のメモリセル対で共用する。なお、メモリセルアレイMC2及び並列多数判定回路20には、それぞれ上記タイミング信号 ϕ_1 等が共通的に入力されているが、説明の簡略化のために図11(b)では図示を省略している。

20

【0102】

以上の構成において、それぞれのニューラルネットワーク回路CCS'を構成するメモリセルブロック15のメモリセル10からは、上記重み付け係数Wが、図9(b)に示す各メモリセル10及び各メモリセルブロック15と同様のタイミングで並列多数判定回路20に出力される。一方、それぞれのニューラルネットワーク回路CCS'を構成するメモリセルブロック15のメモリセル11からは、上記接続有無情報Cが、図9(b)に示す各メモリセル11及び各メモリセルブロック15と同様のタイミングで並列多数判定回路20に出力される。

30

【0103】

そして並列多数判定回路20は、メモリセルアレイMC2から出力されてくる重み付け係数W及び接続有無情報Cと、それらに対応する入力データIと、に基づき、接続有無情報Cが「接続あり」を示している重み付け係数W及び入力データIを用いた上記排他的論理和の演算処理、その演算結果に基づく値「1」の演算結果及び値「0」の演算結果それぞれの数の加算処理、その加算結果に基づく上記合計数の比較処理(図10(b)参照)、及びその比較結果に基づく出力データOの生成処理を、メモリセルアレイMC2における横一行(m個)についてそれぞれ行う。また並列多数判定回路20は、上記横一行についての演算処理、加算処理、比較処理及び生成処理を、各入力データIについて、メモリセルブロック15ごとにシリアル形式で実行し、それぞれの実行結果としての出力データOをシリアル形式で出力する。ここで、対応する接続有無情報Cが「接続なし」を示している場合、並列多数判定回路20は上記演算処理、加算処理、比較処理及び生成処理を行わない。

40

【0104】

以上説明したように、図11(a)においてハッチングで示されるm個のニューロンNRに対してn個のニューロンNRからそれぞれ出力データOが出力され、これらにより当該m個のニューロンNRから1ビットの出力データOがシリアル形式で出力されるニューラルネットワークS2'が、図11(b)に示す構成を備えるニューラルネットワーク集積回路C2'によりモデル化される。

【0105】

50

より具体的に、ニューラルネットワーク集積回路C 2'の構成によれば、出力データO及び並列多数判定回路20の数がそれぞれ一であり、入力データI、メモリセル10及びメモリセル11それぞれの数が、対応すべき脳機能に基づいた既定の二以上の数とされており、重み付け係数W及び接続有無情報Cのそれぞれが脳機能に対応した値とされている。そして並列多数判定回路20が、各入力データIにそれぞれ対応する重み付け係数Wと当該各入力データIとの排他的論理和を入力データIごとにそれぞれ生成し、更に、(値「1」である排他的論理和の総数 - 値「0」である排他的論理和の総数 > 多数判定閾値)であるとき値「1」を出力データOとして出力し、(値「1」である排他的論理和の総数 - 値「0」である排他的論理和の総数 < 多数判定閾値)であるとき値「0」を出力データOとして出力する。よって、各メモリセル10、各メモリセル11及び一の並列多数判定回路20により入力データIごとに乗算処理を実行し、更に並列多数判定回路20により入力データIごとに加算処理及び活性化処理を実行するので、回路規模及びそれに対応するコストを大幅に縮小しつつ、いわゆる並列入力 - 直列出力型のニューラルネットワークS 2'を効率的に実現することができる。

10

【0106】

(D) 実施形態に係るニューラルネットワーク集積回路の第3例について

次に、実施形態に係るニューラルネットワーク集積回路の第3例について、図12を用いて説明する。なお図12において、図8及び図10を用いてそれぞれ説明した実施形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

20

【0107】

図12を用いて説明する実施形態に係るニューラルネットワーク集積回路の第3例は、図9を用いて説明した実施形態に係るニューラルネットワーク集積回路C 1'と、図11を用いて説明した実施形態に係るニューラルネットワーク集積回路C 2'と、を組み合わせた集積回路である。ここで、上記ニューラルネットワーク集積回路C 1'は上述した通り直列入力 - 並列出力型の一段ニューラルネットワークS 1'をモデル化したニューラルネットワーク回路である。一方上記ニューラルネットワーク集積回路C 2'は、上述した通り並列入力 - 直列出力型の一段ニューラルネットワークS 2'をモデル化したニューラルネットワーク回路である。そしてこれらを組み合わせた実施形態に係るニューラルネットワーク集積回路の第3例は、全体として直列入力 - 並列処理 - 直列出力型の多段ニューラルネットワークをモデル化したニューラルネットワーク集積回路であり、更に多くのニューロンNRからなる複雑なニューラルネットワークをモデル化するためのものである。

30

【0108】

先ず、実施形態に係るニューラルネットワーク集積回路の第3例によりモデル化されるニューラルネットワークについて、図12(a)を用いて説明する。図12(a)に示す当該ニューラルネットワークS 1-2は、図12(a)において45度のハッチングで示されるn個のニューロンNRのそれぞれに対してm個のニューロンNRからそれぞれ1ビットの出力データOがシリアル形式で出力され、当該45度のハッチングで示されるニューロンNRと図11(a)において135度のハッチングで示されるm個のニューロンNRとの間で出力データO及び入力データIの授受が行われ、結果的に135度のハッチングで示されるm個のニューロンNRからそれぞれ出力データOがシリアル形式で出力されるニューラルネットワークである。なお上記ニューラルネットワークS 1-2は、全体として、図4を用いて説明したニューラルネットワークS 1を複数並べたニューラルネットワークに相当する。

40

【0109】

上記ニューラルネットワークS 1-2をモデル化した実施形態に係るニューラルネットワーク集積回路の第3例は、図12(b)に示すニューラルネットワーク集積回路C 1-2となる。当該ニューラルネットワーク集積回路C 1-2は、図9を用いて説明したニューラルネットワーク集積回路C 1'の各出力データO(並列的に出力される各出力データO)を、図11を用いて説明したニューラルネットワーク集積回路C 2'における並列多

50

数判定回路20への入力データ(即ち図11(b)に示す入力データI)とし、これにより、当該並列多数判定回路20から上記出力データOをシリアル形式で出力する構成を備える。このように、上記ニューラルネットワーク集積回路C1'とニューラルネットワーク集積回路C2'とを組み合わせることにより、結果的に、図9(a)に例示するニューラルネットワークS1'と、図11(a)に例示するニューラルネットワークS2'と、が組み合わせられた上記ニューラルネットワークS1-2がモデル化される。なお、ニューラルネットワークS1-2に含まれる上記ニューラルネットワーク集積回路C1'及びニューラルネットワーク集積回路C2'それぞれの動作は、図9及び図11を用いてそれぞれ説明した動作と同様となる。なお図12(b)に示すニューラルネットワーク集積回路C1-2では、並列多数判定回路20に対応する直列多数判定回路16が、破線で示される一組の多数判定入力回路12及び直列多数判定回路13により、それぞれ構成されていることになる。

10

【0110】

以上説明したように、図12(a)に示すニューラルネットワークS1-2が、図12(b)に示す直列入力-並列処理-直列出力型の構成を備えるニューラルネットワーク集積回路C1-2によりモデル化される。

【0111】

より具体的に、ニューラルネットワーク集積回路C1-2の構成によれば、ニューラルネットワーク集積回路C1'とニューラルネットワーク集積回路C2'とを備え、一の入力データIが入力されるニューラルネットワーク集積回路C1'から出力される各出力データOがそれぞれ、ニューラルネットワーク集積回路C2'における各入力データIとされており、ニューラルネットワーク集積回路C2'から一の出力データOが出力される。よって、ニューラルネットワーク集積回路C1'とニューラルネットワーク集積回路C2'とを接続することで、回路規模及びそれに対応するコストを大幅に縮小しつつ、多様な脳機能に対応可能なニューラルネットワークS1-2を効率的に実現することができる。

20

【0112】

(E) 実施形態に係るニューラルネットワーク集積回路の第4例について

次に、実施形態に係るニューラルネットワーク集積回路の第4例について、図13及び図14を用いて説明する。なお図13及び図14において、図9及び図11並びに図12を用いてそれぞれ説明した実施形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

30

【0113】

図13を用いて説明する実施形態に係るニューラルネットワーク集積回路の第4例は、図13(a)に示すように、図12を用いて説明した実施形態に係るニューラルネットワーク集積回路C1-2において、それを構成する上記ニューラルネットワーク集積回路C1'と上記ニューラルネットワーク集積回路C2'との間にパイプラインレジスタ21を介在させた構成を備えたニューラルネットワーク集積回路C1-3である。このパイプラインレジスタ21が、本発明に係る「レジスタ部」の一例に相当する。そしてパイプラインレジスタ21は、メモリセルアレイMC1のビット幅に相当する数のデータを一時的に記憶すると共に、外部からのイネーブル信号ENにより、その出力動作が制御される。このイネーブル信号ENは、予め設定された基準クロック信号のうちの偶数基準クロックに相当するタイミング信号である。そしてニューラルネットワーク集積回路C1-3は、図13(b)に示すように全体として、ニューラルネットワーク集積回路C1'におけるメモリセルアレイMC1と、ニューラルネットワーク集積回路C2'におけるメモリセルアレイMC2と、の間に、1ビットの入力データIが例えばm個シリアル形式で入力されると共に上記イネーブル信号ENが入力され、これらに対応した1ビットの出力データOが例えばm個シリアル形式で出力される並列演算器PPを介在させた構成を備える。このとき、メモリセルアレイMC1及びメモリセルアレイMC2はそれぞれ、例えば256ビット幅で512ワード(Word)分の規模を備えており、アドレス指定用の例えば8ビットのアドレスデータADがそれぞれに入力される。そしてこの場合の並列演算器PPは、25

40

50

6ビット分の多数判定入力回路12及び直列多数判定回路13と、上記パイプラインレジスタ21と、256ビットに対応する並列多数判定回路20と、により構成される。

【0114】

以上の構成において、ニューラルネットワークS1-3に含まれる上記ニューラルネットワーク集積回路C1'及びニューラルネットワーク集積回路C2'それぞれの動作は、図9及び図11を用いて説明した動作と同様となる。一方パイプラインレジスタ21は、例えば、ニューラルネットワーク集積回路C2'のメモリセルアレイMC2から読み出した重み付け係数W及び接続有無情報Cに基づいて並列多数判定回路20において出力データOの生成/出力処理を行っているタイミングでニューラルネットワーク集積回路C1'のメモリセルアレイMC1から読み出した出力データOを一時的に記憶する。そして、上記重み付け係数W及び接続有無情報Cに基づく並列多数判定回路20の処理が完了したタイミングで、メモリセルアレイMC1から読み出して記憶していた出力データOを並列多数判定回路20に出力してそれに基づいた出力データOの生成/出力処理を行わせる。この処理により、見かけ上はメモリセルアレイMC1からの出力データOの読み出しと、メモリセルアレイMC2からの重み付け係数W及び接続有無情報Cの読み出しと、を同時に行わせることができ、結果的に、図12を用いて説明したニューラルネットワークS1-2に対して略二倍の処理速度を実現させることができる。

10

【0115】

次に、図13に示すニューラルネットワーク集積回路C1-3における特に並列演算器PPの細部構成について、図14を用いて説明する。

20

【0116】

先ず図14(a)に示すように並列演算器PPは、メモリセルアレイMC1のビット幅に相当する数の上記多数判定入力回路12及び上記直列多数判定回路13からなる直列多数判定回路16と、メモリセルアレイMC1のビット幅に相当する上記パイプラインレジスタ21と、出力フリップフロップ回路22を介して出力データOを出力する上記並列多数判定回路20と、により構成されている。この構成においてパイプラインレジスタ21は図14(a)に示すように、メモリセルアレイMC1のビット幅に相当する出力レジスタ21U及び入力レジスタ21Lにより構成されており、上記イネーブル信号ENが入力レジスタ21Lに入力される。そして入力レジスタ21Lは、イネーブル信号ENが入力されるタイミングでそれに記憶(ラッチ)されているデータを並列多数判定回路20に出力すると共に、当該タイミングで出力レジスタ21Uに記憶されているデータを引き出して(即ちシフトさせて)記憶(ラッチ)する。またこれにより出力レジスタ21Uは、入力レジスタ21Lによりそのデータが引き出されたタイミングで、次の出力データOを記憶(ラッチ)する。以上の入力レジスタ21L及び出力レジスタ21Uの動作が繰り返されることにより、上述したパイプラインレジスタ21としての動作が実現される。

30

【0117】

次に、上記多数判定入力回路12及び直列多数判定回路13の細部構成について、図14(b)を用いて説明する。図14(b)に示すように一の直列多数判定回路16内の多数判定入力回路12は、排他的否定論理和回路12Aと、マスクフリップフロップ回路12Bと、により構成されている。この構成において排他的否定論理和回路12Aには、メモリセルアレイMC1からの重み付け係数Wと、1ビットの入力データIと、が入力され、これらの排他的否定論理和の結果を上記出力データEとして直列多数決回路13に出力する。またマスクフリップフロップ回路12Bは、メモリセルアレイMC1からの接続有無情報Cと、上記イネーブル信号ENと、が入力され、イネーブル信号ENが入力されたタイミングで上記接続有無情報Cを直列多数決回路13に出力する。そして直列多数決回路13は、上記出力データE及び上記接続有無情報Cに基づいた上述した動作により出力データOを生成して、パイプラインレジスタ21の出力レジスタ21Uに出力する。このとき、上記既定の多数判定閾値を直列多数決回路13内の図示しないレジスタ内に保持してそれを参照することで、上述した直列多数決回路13としての動作を実現できる。

40

【0118】

50

次に、上記並列多数判定回路20の細部構成について、図14(c)を用いて説明する。図14(c)に示すように並列多数判定回路20は、入力データIの数と同数の排他的否定論理和回路20A及びマスクフリップフロップ回路20Bと、並列多数決回路20Cと、により構成されている。この構成において排他的否定論理和回路20Aには、メモリセルアレイMC2からの1ビットの重み付け係数Wと、パイプラインレジスタ21の入力レジスタ21Lからの1ビットの出力データOと、が入力され、これらの排他的論理和の結果を並列多数決回路20Cに出力する。またマスクフリップフロップ回路20Bは、メモリセルアレイMC2からの接続有無情報Cと、上記イネーブル信号ENと、が入力され、イネーブル信号ENが入力されたタイミングで上記接続有無情報Cを並列多数決回路20Cに出力する。そして並列多数決回路20Cは、メモリセルアレイMC2からの一組の重み付け係数W及び接続有無情報Cに対応する排他的否定論理和回路12A及びマスクフリップフロップ回路20Bそれぞれからの出力に基づいた上述した動作をメモリセルアレイMC1からの出力データOの数(図13及び図14に例示する場合は256)だけ繰り返し、出力フリップフロップ回路22を介してシリアル形式で出力データOとして出力する。このとき、上記既定の多数判定閾値を並列多数決回路20C内の図示しないレジスタ内に保持してそれを参照することで、上述した並列多数決回路20としての動作を実現できる。

10

【0119】

このとき、上述したパイプラインレジスタ21の動作により並列演算器PPでは、例えば図14(d)に示すように、メモリセルアレイMC1からの256ビット分の出力データOに対する処理(図14(d)において「メモリセルブロック15_{U1}」と示す)が終了すると、次にメモリセルアレイMC1からの次の256ビット分の出力データOに対する処理(図14(d)において「メモリセルブロック15_{U2}」と示す)と、メモリセルアレイMC2からの256ビット分の重み付け係数W及び接続有無情報Cに対する処理(図14(d)において「メモリセルブロック15_{L1}」と示す)と、が、見かけ上同時並列的に実行される。そして、メモリセルブロック15_{U2}に対応する出力データO並びにメモリセルブロック15_{L1}に対応する重み付け係数W及び接続有無情報Cに対する処理が終了すると、次にメモリセルアレイMC1からの更に次の256ビット分の出力データOに対する処理(図14(d)において「メモリセルブロック15_{U3}」と示す)と、メモリセルアレイMC2からの次の256ビット分の重み付け係数W及び接続有無情報Cに対する処理(図14(d)において「メモリセルブロック15_{L2}」と示す)と、が、見かけ上同時並列的に実行される。以降は、メモリセルアレイMC1及びメモリセルアレイMC2それぞれからの256ビット分の出力データO並びに重み付け係数W及び接続有無情報Cに対して、逐次的且つ同時並列的な処理が実行される。

20

30

【0120】

なお、図14(b)に示す多数判定入力回路12及び直列多数判定回路13の細部構成、並びに図14(c)に示す並列多数判定回路20の細部構成は、図8以降に示す各メモリセル11からの上記接続有無情報Cの出力タイミングが、図8以降に示す各メモリセル10からの上記重み付け係数Wの出力タイミングよりも例えば一サイクル早いことを前提とした構成である。この出力タイミングのずれを吸収するのが、図14(b)及び図14(c)にそれぞれ示すマスクフリップフロップ回路12B及びマスクフリップフロップ回路20Bの機能である。一方、上記重み付け係数Wの出力タイミングと上記接続有無情報Cの出力タイミングとを同時並行的とすることも可能である。そしてこの場合、図14(b)及び図14(c)にそれぞれ示すマスクフリップフロップ回路12B及びマスクフリップフロップ回路20Bは、多数判定入力回路12及び並列多数判定回路20としては不要となる。

40

【0121】

以上説明したように、図13及び図14に示すニューラルネットワーク集積回路C1-3によれば、図12(a)に示すニューラルネットワークS1-2を、約二倍の処理速度をもってモデル化することができる。なお図14を用いて説明した直列多数判定回路16

50

の細部構成は、図 1 2 を用いて説明したニューラルネットワーク集積回路 C 1 - 2 に含まれる直列多数判定回路 1 6 の細部構成として適用することもできる。

【 0 1 2 2 】

より具体的に、ニューラルネットワーク集積回路 C 1 - 3 によれば、ニューラルネットワーク集積回路 C 1 ' とニューラルネットワーク集積回路 C 2 ' との間に備えられたパイプラインレジスタ 2 1 により、ニューラルネットワーク集積回路 C 1 ' から出力される各出力データ O を一基準クロック分だけバッファリングして、ニューラルネットワーク集積回路 C 2 ' における各入力データ I として当該ニューラルネットワーク集積回路 C 2 ' に出力する。よって、ニューラルネットワーク集積回路 C 2 ' における並列多数判定回路 2 0 における見かけ上の並列処理が可能となり、多様な脳機能に対応可能なニューラルネットワーク S 1 - 2 を効率的且つ高速化して実現することができる。

10

【 0 1 2 3 】

(F) 実施形態に係るニューラルネットワーク集積回路の第 5 例について

次に、実施形態に係るニューラルネットワーク集積回路の第 5 例について、図 1 5 を用いて説明する。なお図 1 5 において、図 1 3 及び図 1 4 を用いてそれぞれ説明した実施形態に係るニューラルネットワーク回路と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

【 0 1 2 4 】

図 1 5 を用いて説明する実施形態に係るニューラルネットワーク集積回路の第 5 例は、図 1 5 (a) に示すように、図 1 3 及び図 1 4 を用いて説明した実施形態に係るニューラルネットワーク集積回路 C 1 - 3 を例えば四つ備え、これらの動作を制御部 C T により制御する構成を備えたニューラルネットワーク集積回路 C P である。より具体的にニューラルネットワーク集積回路 C P は、図 1 5 (a) に示すように、メモリブロック M A 上に集積された四つのニューラルネットワーク集積回路 C 1 - 3 と、各ニューラルネットワーク集積回路 C 1 - 3 に対応した四つの入力セクタ I S と、入力制御用で F I F O (First In First Out) 形式のメモリ M 1 と、出力制御用で F I F O 形式のメモリ M 2 と、制御部 C T と、バス B と、により構成されている。

20

【 0 1 2 5 】

この構成において 1 ビットの入力データ I は、書き込みストロープ W S により制御されるメモリ M 1 を介してバス B にシリアル形式で出力される。これにより各ニューラルネットワーク集積回路 C 1 - 3 のそれぞれは、対応する入力セクタ I S を介して入力データ I をシリアル形式で読み込み、上述してきた処理を施し、1 ビットの出力データ O を生成してシリアル形式でバス B に出力する。そして、各ニューラルネットワーク集積回路 C 1 - 3 からの出力データ O は、入力セクタ I S を介してメモリ M 2 に読み込まれる。その後、当該読み込まれた出力データ O は、読み出しストロープ R S により制御されるメモリ M 2 を介して外部に出力される。このとき制御部 C T は、外部からの実行信号 R N に基づいて、各ニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 及びメモリセルアレイ M C 2 にそれぞれ送信されるアドレスデータ A D 及びイネーブル信号 E N (図 1 3 参照) と、スタート信号 S T と、を生成して各ニューラルネットワーク集積回路 C 1 - 3 に出力する。

30

40

【 0 1 2 6 】

次に、上記制御部 C T の細部構成について、図 1 5 (b) を用いて説明する。

【 0 1 2 7 】

図 1 5 (b) に示すように制御部 C T は、アップカウンタ 2 5 ₁ 乃至アップカウンタ 2 5 ₅ と、サイクリックカウンタ 2 6 と、アップカウンタ 2 5 ₁ 乃至アップカウンタ 2 5 ₅ にそれぞれ対応するレジスタ 2 7 ₁ 乃至レジスタ 2 7 ₅ と、サイクリックカウンタ 2 6 に対応するレジスタ 2 7 _m と、により構成されている。

【 0 1 2 8 】

この構成においてサイクリックカウンタ 2 6 は、上記実行信号 R N とレジスタ 2 7 _m に記憶されているカウント値に基づいて上記スタート信号 S T を生成し、各ニューラルネッ

50

トワーク集積回路 C 1 - 3 及びアップカウンタ 2 5₁ 乃至アップカウンタ 2 5₅ にそれぞれ出力する。これによりアップカウンタ 2 5₁ は、図示しない F I F O 型の入力メモリと最初のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 宛のアドレスデータ A D 及びイネーブル信号 E N を生成し、当該入力メモリと最初のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 に出力する。次にアップカウンタ 2 5₂ は、最初のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 2 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 宛のアドレスデータ A D 及びイネーブル信号 E N を生成し、当該最初のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 2 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 に出力する。次にアップカウンタ 2 5₃ は、第 2 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 3 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 宛のアドレスデータ A D 及びイネーブル信号 E N を生成し、当該第 2 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 3 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 に出力する。次にアップカウンタ 2 5₄ は、第 3 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 4 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 宛のアドレスデータ A D 及びイネーブル信号 E N を生成し、当該第 3 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び第 4 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 1 に出力する。最後にアップカウンタ 2 5₅ は、第 4 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び図示しない F I F O 型の出力メモリ宛のアドレスデータ A D 及びイネーブル信号 E N を生成し、当該第 4 のニューラルネットワーク集積回路 C 1 - 3 のメモリセルアレイ M C 2 及び出力メモリに出力する。これらにより各ニューラルネットワーク集積回路 C 1 - 3 は、上記スタート信号 S T、上記アドレスデータ A D 及びイネーブル信号 E N に基づき、例えば図 1 4 (d) に例示する態様で上述した動作をシリアル形式で実行し、最終的に出力データ O がメモリ M 2 を介して出力される。なお制御部 C T は、モデル化すべきニューラルネットワークの態様 (例えば、ニューロン N R の数やそれらの間の接続態様等) を変更すべき場合、上記アドレス A D 及びイネーブル信号 E N の内容又は出力先を変更することで、当該態様の変更に対応させてニューラルネットワーク集積回路 C P の動作を変更する。

10

20

30

【 0 1 2 9 】

以上説明したように、図 1 5 に示すニューラルネットワーク集積回路 C P によれば、更に大規模なニューラルネットワークを、その構成の自由度を高めつつモデル化することができる。

【 0 1 3 0 】

(G) 実施形態に係るニューラルネットワーク集積回路の第 6 例等について

次に、実施形態に係るニューラルネットワーク集積回路の第 6 例及び第 7 例について、図 1 6 を用いて説明する。なお図 1 6 において、図 1 2 を用いて説明した実施形態に係るニューラルネットワーク回路 C 1 - 2 と同様の構成部材については、同様の部材番号を付して細部の説明を省略する。

40

【 0 1 3 1 】

先ず、実施形態に係るニューラルネットワーク集積回路の第 6 例を、図 1 6 (a) を用いて説明する。当該第 6 例は図 1 6 (a) に示すように、図 1 2 を用いて説明した実施形態に係るニューラルネットワーク集積回路 C 1 - 2 を例えば三つ備え、これらが直列的に接続された構成を備えるニューラルネットワーク集積回路である。このときニューラルネットワーク集積回路 C 1 - 2 は、上述した通り直列入力 - 並列処理 - 直列出力型の多段ニューラルネットワークをモデル化したニューラルネットワーク集積回路である。また図 1 6 (a) に示す構成は、図 5 を用いて説明した関連形態に対応する態様で上記ニューラルネットワーク集積回路 C 1 - 2 を直列的に接続したニューラルネットワーク集積回路である。よってこの場合、一つのニューラルネットワーク集積回路 C 1 - 2 の出力データ O の

50

数（シリアル形式で出力される出力データOの数）は、次段に接続されるニューラルネットワーク集積回路C1-2の入力データIの数（シリアル形式で出力される入力データIの数）に相当している必要がある。

【0132】

次に、実施形態に係るニューラルネットワーク集積回路の第7例を、図16(b)を用いて説明する。当該第7例は図16(b)に示すように、図12を用いて説明した実施形態に係るニューラルネットワーク集積回路C1-2を例えば三つ備え、共通の入力データIによりこれらが並列的に接続された構成を備えるニューラルネットワーク集積回路である。この図16(b)に示す構成は、図6を用いて説明した関連形態に対応する態様で上記ニューラルネットワーク集積回路C1-2を並列的に接続したニューラルネットワーク集積回路である。よってこの場合、図16(a)に示したニューラルネットワーク集積回路の場合のような出力データO及び入力データIそれぞれにおけるビット数についての制限はない。

10

【0133】

以上説明したように、図16に示す各ニューラルネットワーク集積回路によれば、更に大規模なニューラルネットワークをモデル化することができる。

【0134】

より具体的に、図16(a)に示す各ニューラルネットワーク集積回路によれば、複数のニューラルネットワーク集積回路C1-2が直列に接続されており、一のニューラルネットワーク集積回路C1-2からの一の出力データOが、直後段のニューラルネットワーク集積回路C1-2への一の入力データIとされているので、より多様な脳機能に対応可能なニューラルネットワークを効率的に実現することができる。また図16(b)に示す各ニューラルネットワーク集積回路によれば、複数のニューラルネットワーク集積回路C1-2のそれぞれに対して一の入力データIが共通に入力されており、当該複数のニューラルネットワーク集積回路C1-2からそれぞれ別個に出力データOがずつ出力されるので、より多様な脳機能に対応可能なニューラルネットワークを効率的に実現することができる。

20

【0135】

以上それぞれ説明したように、実施形態に係るニューラルネットワーク回路CCS'及びニューラルネットワーク集積回路C1-2等の構成及び動作によれば、関連形態に係るニューラルネットワーク回路CS及びニューラルネットワーク集積回路C1等の構成及び動作と同様に、回路規模及びそれに対応するコストを大幅に縮小しつつ、ニューラルネットワーク回路を効率的に実現することができる。

30

【0136】

(III) 変形形態

次に、本発明に係る上述した関連形態及び実施形態を変形した変形形態について、図17及び図18を用いて説明する。なお、図17は変形形態に係るニューラルネットワークを説明する図等であり、図18は変形形態に係るニューラルネットワーク集積回路の細部構成を示す図である。

【0137】

以下に説明する変形形態は、上記関連形態及び上記実施形態としてそのモデル化を説明してきたニューラルネットワークに対して、脳機能をより一般的にモデル化する手法として、新たに「傾向係数」の概念を追加するものである。なお以下の説明において、特に図1及び図14を用いて説明した構成に含まれる部材等と同じ部材等については、同様の部材等の番号を付して細部の説明は省略する。

40

【0138】

先ず、上記傾向係数の概念を追加してモデル化したニューラルネットワークについて図17を用いて説明する。

【0139】

上述した関連形態及び実施形態において説明したニューラルネットワークのニューロン

50

NRでは、外部から入力される複数の入力データIのそれぞれに対する上記乗算処理、上記加算処理及び上記活性化処理を当該ニューロンNRで実行し、その結果を出力データOとしていた。これに対し、変形形態に係るニューラルネットワークでは、各入力データIに対する上記乗算処理及び上記加算処理の結果に対して、当該ニューラルネットワークによりモデル化される脳機能に基づいて予め設定された、当該脳機能としての傾向をニューロンごとに示す上記傾向係数を更に加算する。そして変形形態に係るニューラルネットワークでは、当該傾向係数の加算結果に対して上記活性化処理を適用し、その結果を、変形形態に係る出力データとして出力する。即ち変形形態に係るニューラルネットワークの場合、「加算処理」とは、上述した関連形態及び実施形態それぞれに係るニューラルネットワークにおける加算処理に対して上記傾向係数の加算処理を更に加えることをいう。この変形形態に係る傾向係数は、上述したように脳機能としての傾向をニューロンごとに示し且つ予め設定された係数であり、その絶対値は、当該ニューロンに入力される入力データの数n以下の任意の整数である。ここで、上記「傾向」とは、例えば、モデル化の対象である脳機能においてニューロンごとの調整機能としてその存在が認められる、脳機能等の傾向をいう。更に上記傾向係数は、一般には「バイアス値 (Bias value)」と称される場合もある。この傾向係数としての具体的な値は、そのニューロンについての上記傾向に対応して、上記入力データIの数以下の整数として、ニューロンごとに予め設定されている。より具体的に例えば、一つのニューロンに対して三つの入力データIが入力されるとすると、上記傾向係数として取り得る値は、「-3」乃至「3」の七つ（「0」を含む）である。ここで、傾向係数が「0」の場合が、上述してきた関連形態及び実施形態に係るニューラルネットワークであることになる（図1(a)参照）。

【0140】

そして、上述した変形形態に係るニューラルネットワークとしての一連の乗算処理、加算処理（傾向係数の加算処理を含む。以下、変形形態において同じ。）及び活性化処理を数式で表すと、図17(a)に示す式(2)となる。このとき、関連形態及び実施形態の場合と同様に各重み付け係数Wを各入力データIにそれぞれ乗算する乗算処理が、変形形態に係るニューロンNR間の電気信号のやり取りにおけるシナプスの作用に相当する。そして図17(a)に例示するように、変形形態に係る加算処理後の値全体に対して活性化関数Fを適用して出力データOを出力することが、変形形態に係る活性化処理に相当する。

【0141】

なお、入力データIと重み付け係数Wとの各乗算結果の加算結果に対する変形形態に係る傾向係数 B_n の更なる加算処理を回路的に実現する場合、以下に説明する変形形態では、「1」を分割単位として当該傾向係数 B_n を分割（分解）し、上記各乗算結果のそれぞれに対して「1」を加算又は減算する処理を、傾向係数 B_n の値に等しい数の乗算結果について行う。即ち、例えば傾向係数 B_n の値が「2」である場合、図17(a)に示す例えば入力データ I_1 と重み付け係数 W_1 の乗算結果に「1」を加算すると共に、入力データ I_2 と重み付け係数 W_2 の乗算結果にも「1」を加算し、これらを含めて活性化関数Fを適用させる。また、例えば傾向係数 B_n の値が「-3」である場合、図17(a)に示す例えば入力データ I_1 と重み付け係数 W_1 の乗算結果及び入力データ I_2 と重み付け係数 W_2 の乗算結果からそれぞれ「1」を減算し、更に入力データ I_3 と重み付け係数 W_3 の乗算結果からも「1」を減算し、これらを含めて活性化関数Fを適用させる。このとき、いずれの乗算結果に対して「1」の加算又は減算を行うかは、モデル化される脳機能に対応して予め設定されている。

【0142】

ここで変形形態に係る加算処理等を回路的に実現する場合、上記接続有無情報Cの意味（定義）が、上記関連形態及び上記実施形態とは異なってくる。即ち図17(b)に示すように、変形形態に係る傾向係数 B_n の値が「0」である場合は、上記関連形態及び上記実施形態の構成に相当する。そしてこの場合は、変形形態に係るニューロンNR間であっても、当該ニューロンNR間について接続有無情報Cが「接続あり」を示している（

図17(b)において「1」とされている)場合のみ、対応する入力データIと重み付け係数Wとの乗算結果が一方のニューロンNR Rに入力される(図17(b)最上段及び上から二段目参照)。このとき、上記活性化処理後の出力データの値は、「0」又は「1」となる。そしてその後、そのニューロンNR Rにおける関連形態及び実施形態と同様の加算処理及び活性化処理に供される。

【0143】

これに対し、変形形態に係る傾向係数 B_n が「0」でない場合、当該傾向係数 B_n を分割した結果を加算又は減算する対象の乗算結果(入力データIと重み付け係数Wとの乗算結果)について、接続有無情報Cが「1」とされている(即ち「接続あり」を示している)場合には、正の値である傾向係数 B_n を分解した値としての「1」が当該乗算結果に「加算」されて、一方のニューロンNR Rに入力される(図17(b)最下段参照)。一方、接続有無情報Cが「0」とされている(即ち「接続なし」を示している)場合には、負の値である傾向係数 B_n を分解した値としての「-1」が上記乗算結果に加算(即ち「1」が減算)されて、一方のニューロンNR Rに入力される(図17(b)下から二段目参照)。この場合、上記活性化処理後の出力データの値は、「-2」又は「0」(傾向係数 B_n が負の値である場合。図17(b)下から二段目参照。)、或いは、「2」又は「0」(傾向係数 B_n が正の値である場合。図17(b)最下段参照。)となる。そしてその後、そのニューロンNR Rにおける変形形態としての加算処理及び活性化処理に供される。

10

【0144】

次に、図13に示すニューラルネットワーク集積回路C1-3に対して変形形態を適用した場合の、特に並列演算器(図13符号PP参照)の細部構成について、図14を参照しつつ、図18を用いて説明する。

20

【0145】

先ず図18(a)に示すように変形形態に係る並列演算器PPPは、メモリセルアレイMC1のビット幅に相当する数の変形形態に係る多数判定入力回路12'及び上記直列多数判定回路13からなる直列多数判定回路16'と、ニューラルネットワーク集積回路C1-3と同様のパイプラインレジスタ21と、ニューラルネットワーク集積回路C1-3と同様の出力フリップフロップ回路22を介して出力データOを出力する変形形態に係る並列多数判定回路20'と、により構成されている。この構成においてパイプラインレジスタ21は、図14(a)を用いて説明した場合と同様の構成を備え、同様の動作を実現する。また変形形態に係る並列演算器PPPにおいては、重み付け係数Wと接続有無情報Cとが一サイクル異なったタイミングで入力される実施形態に係る並列演算器PPの場合と異なり、重み付け係数W、接続有無情報C及び傾向係数 B_n は、メモリセルアレイMC1又はメモリセルアレイMC2から、それぞれ相互に同時並行的に並列演算器PPPに入力される。

30

【0146】

次に、変形形態に係る多数判定入力回路12'の細部構成について、図18(b)を用いて説明する。図18(b)に示すように一の直列多数判定回路16'内の多数判定入力回路12'は、排他的論理和回路25Aと、セクタ25B及びセクタ25Fと、反転増幅回路25C及び反転増幅回路25Eと、論理積回路25Dと、加算回路25Gと、により構成されている。このとき、重み付け係数W、接続有無情報C及び傾向係数 B_n が同時並行的に並列演算器PPPに入力されるため、図14(b)に例示したマスクフリップフロップ回路12Bは不要となる。以上の構成において排他的論理和回路25Aには、メモリセルアレイMC1からの重み付け係数Wと、1ビットの入力データIと、が入力され、これらの排他的論理和の結果をセクタ25Bの「0」端子に出力する。一方論理積回路25Dには、メモリセルアレイMC1からの接続有無情報Cと、傾向係数 B_n を分割した1ビットのデータを反転増幅回路25Cにより反転したデータと、が入力され、それらの論理積の結果がセクタ25Bの制御端子に出力される。これによりセクタ25Bは、接続有無情報C及び傾向係数 B_n に基づき、上記排他的論理和の結果又は0データのい

40

50

ずれか一方を選択し、加算回路 25G の一方の端子に出力する。次にセレクタ 25F の制御端子には上記傾向係数 B_n を分割した 1 ビットのデータが入力され、更にセレクタ 25F の「1」端子には、接続有無情報 C を反転増幅回路 25E により反転したデータが入力される。これによりセレクタ 25F は、傾向係数 B_n に基づき、上記反転増幅回路 25E からの出力データ又は 0 データのいずれか一方を選択し、加算回路 25G の他方の端子に出力する。これらにより加算回路 25G は、セレクタ 25B 及びセレクタ 25F からの出力データを加算し、出力データ OP_n として直列多数決回路 13 に出力する。このとき当該出力データ OP_n の取り得る値は、図 17 (b) に示すように「-2」、「-1」、「0」、「1」又は「2」の五種類となるため、出力データ OP_n は 3 ビットのデータとなる。そして直列多数決回路 13 は、出力データ OP_n に対して図 14 に例示した場合と同様の活性化処理に相当する処理を行い、その結果を出力データ OO_n として (図 17 (a) 参照)、パイプラインレジスタ 21 の出力レジスタ 21U に出力する。このとき、上記既定の多数判定閾値を直列多数決回路 13 内の図示しないレジスタ内に保持してそれを参照することで、上述した直列多数決回路 13 としての動作を実現できる点は、図 14 に例示した直列多数決回路 13 と同様である。

10

【0147】

次に、変形形態に係る並列多数判定回路 20' の細部構成について、図 18 (c) を用いて説明する。図 18 (c) に示すように並列多数判定回路 20' は、図 18 (b) を用いて構成等を説明した多数判定入力回路 12' と同様の構成等を備え且つメモリセルアレイ MC2 のビット幅に相当する数の多数判定入力回路 12' と、並列多数決回路 30 と、により構成されている。この構成において多数判定入力回路 12' には、メモリセルアレイ MC2 からの 1 ビットの重み付け係数 W 及び接続有無情報 C 並びに傾向係数 B_n と、パイプラインレジスタ 21 の入力レジスタ 21L からの 1 ビットの出力データ O と、が入力される。そして各多数判定入力回路 12' は、重み付け係数 W、接続有無情報 C 並びに傾向係数 B_n 及び入力レジスタ 21L からの 1 ビットの出力データ O に対して図 18 (b) を用いて説明した動作と同様の動作を実行した結果を、並列多数決回路 30 に出力する。そして並列多数決回路 30 は、メモリセルアレイ MC2 からの一組の重み付け係数 W 及び接続有無情報 C 並びに傾向係数 B_n に対応する各多数判定入力回路 12' それぞれからの出力データ OP_n に基づいた上述した多数決動作を、メモリセルアレイ MC1 からの出力データ O の数 (例えば図 13 及び図 14 で例示した場合と同様の 256) だけ繰り返し、出力フリップフロップ回路 22 を介してシリアル形式で出力データ O として出力する。このとき、変形形態に係る並列多数決回路 30 の構成として具体的には、例えば、隣接するビットに対応する出力データ OP_n 同士を加算するための加算回路を階層的に備え、その最終的な加算結果の符号ビットを上記出力フリップフロップ回路 22 に出力する構成が考えられる。また、上記既定の多数判定閾値を並列多数決回路 30 内の図示しないレジスタ内に保持してそれを参照することで、上述した並列多数決回路 30 としての動作を実現できる点は、図 14 に例示した並列多数決回路 20C と同様である。

20

30

【0148】

また、パイプラインレジスタ 21 の動作によりメモリセルアレイ MC1 及びメモリセルアレイ MC2 それぞれからの 256 ビット分の出力データ O、重み付け係数 W 及び接続有無情報 C 並びに傾向係数 B_n に対して、逐次的且つ同時並列的な処理が実行されることは、図 14 に示す並列演算器 PP と同様である。

40

【0149】

以上説明したように、図 17 及び図 18 に示す変形形態によれば、傾向係数 B_n を加味した例えば図 12 (a) に示すニューラルネットワーク S1-2 を、約二倍の処理速度をもってモデル化することができる。

【0150】

更に、変形形態に係るニューラルネットワーク集積回路に含まれる並列演算器 PPP の構成及び動作によれば、変形形態に係るニューラルネットワーク (図 17 参照) をモデル化したニューラルネットワーク集積回路として、回路規模及びそれに対応するコストを大

50

幅に縮小することができる。

【 0 1 5 1 】

より具体的に、変形形態に係るニューラルネットワーク集積回路に含まれる並列演算器 P P P の構成及び動作によれば、脳機能に基づく既定の傾向係数 B_n が各乗算処理の加算結果に対して更に加算されるべき場合、当該傾向係数 B_n の絶対値に等しい数のいずれかの多数判定入力回路 1 2 ' に当該傾向係数 B_n が分けて入力される。そして上記接続有無情報 C に応じて、各乗算処理の加算結果に対する当該傾向係数 B_n の更なる加算結果が多数判定入力回路 1 2 ' から並列多数決回路 3 0 に出力される。これらにより並列多数決回路 3 0 は、結果的に、傾向係数 B_n の更なる加算結果を各多数判定入力回路 1 2 ' について加算した合計値が正又は 0 であるとき値「 1 」を出力データ O として出力し、当該合計値が負であるとき値「 0 」を出力データ O として出力する。よって、脳機能に基づく既定の傾向係数 B_n が導入されるべき場合であっても、回路規模及びそれに対応するコストを大幅に縮小しつつ、人の脳機能に対応したニューラルネットワークを効率的に実現することができる。

10

【 0 1 5 2 】

なお、上述してきた関連形態並びに実施形態及び変形形態においては、出力データ O、入力データ I、重み付け係数 W 等において、それらの値を「 1 」又は「 0 」等とする場合について説明したが、これ以外に、回路設計上の便宜等により、上記値「 0 」を値「 - 1 」として各ニューラルネットワーク回路又はニューラルネットワーク集積回路を構成してもよい。

20

【産業上の利用可能性】

【 0 1 5 3 】

以上それぞれ説明したように、本発明はニューラルネットワークをモデル化したニューラルネットワーク回路等の分野に利用することが可能であり、特に、製造コストの低減や効率的なニューラルネットワーク回路等を開発する場合に適用すれば、特に顕著な効果が得られる。

【符号の説明】

【 0 1 5 4 】

- 1、 1 0、 1 1、 メモリセル
- 2 多数判定回路
- 1 2、 1 2 ' 多数判定入力回路
- 1 2 A、 2 0 A 排他的否定論理和回路
- 1 2 B、 2 0 B マスクフリップフロップ回路
- 1 3、 1 6、 1 6 ' 直列多数判定回路
- 1 5、 1 5_{U1}、 1 5_{U2}、 1 5_{L1}、 1 5_{U3}、 1 5_{L2} メモリセルブロック
- 2 0、 2 0 ' 並列多数判定回路
- 2 0 C、 3 0 並列多数決回路
- 2 1 パイプラインレジスタ
- 2 1 U 出力レジスタ
- 2 1 L 入力レジスタ
- 2 2 出力フリップフロップ回路
- 2 5₁、 2 5₁、 2 5₁、 2 5₁、 2 5₅ アップカウンタ
- 2 5 A 排他的論理和回路
- 2 5 B、 2 5 F セレクタ
- 2 5 C、 2 5 E 反転増幅回路
- 2 5 D 論理積回路
- 2 5 G 加算回路
- 2 6 サイクリックカウンタ
- 2 7₁、 2 7₂、 2 7₃、 2 7₄、 2 7₅、 2 7_m レジスタ
- I、 I₁、 I₂、 I₃、 I₄、 I₅、 I₆、 I₇、 I₈、 I_n、 I_m 入力データ

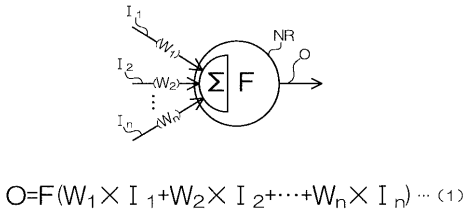
30

40

50

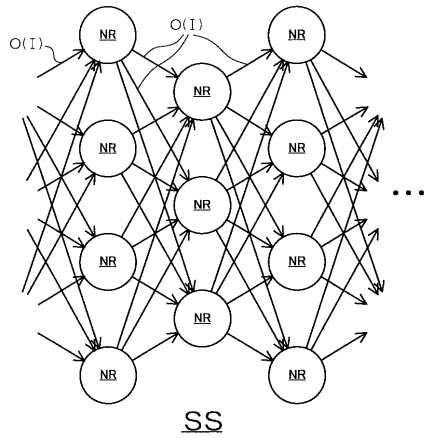
NR、NRR	ニューロン	
G ₁	XNORゲート	
G ₂	XORゲート	
B _n	傾向係数	
O、O ₁ 、O ₂ 、O ₃ 、O ₄ 、O ₅ 、O _n 、O _m 、O _{2m} 、O ₁₀ 、O ₂₀ 、O ₃₀ 、 O ₄₀ 、O ₁₁ 、O ₂₁ 、O ₃₁ 、O ₄₁ 、O _{1n} 、O _{2n} 、O _{3n} 、O _{n0} 、O _{n1} 、 OO、OP _n 、E ₁ 、E ₂ 、E ₃ 、E ₄ 、E ₁₀ 、E ₂₀ 、E ₃₀ 、E ₄₀ 、E ₁₁ 、E ₂₁ 、E ₃₁ 、E ₄₁ 、E _{n0} 、E _{n1}	出力データ	
W ₁ 、W ₂ 、W _n	重み付け係数	
S、S ₁ 、SS ₁ 、SS ₂ 、SS ₃ 、S ₁ '、S ₂ '、S ₁₋₂ 、S ₁₋₃	ニューラルネットワーク	10
S'	ネットワーク	
C ₁ 、C ₂ 、C ₃ 、C ₄ 、CC ₁ 、CC ₂ 、CC ₃ 、C ₁ '、C ₂ '、CP、C ₁₋₂ 、 C ₁₋₃	ニューラルネットワーク集積回路	
CS、CS ₁ 、CS ₂ 、CS ₃ 、CS ₄ 、CS ₅ 、CS ₁₁ 、CS ₁₂ 、CS ₂₁ 、C ₂₂ 、 CS ₃₁ 、CS ₃₂ 、CS _{1n} 、CS _{1m} 、CS _{2n} 、CS _{2m} 、CS _{3n} 、CS _{3p} 、CS _{4q} 、CS _m 、CCS'	ニューラルネットワーク回路	
CS'	ネットワーク回路	
T ₁ 、T ₂ 、T ₃ 、T ₄ 、T ₅ 、T ₆ 、T ₇ 、T ₈ 、T ₉ 、T ₁₀ 、T ₁₁ 、T ₁₂ 、 T ₁₃ 、T ₁₄ 、T ₁₅ 、T ₁₆ 、T ₁₇ 、T ₁₈ 、T ₁₉ 、T ₂₀ 、T ₂₁ 、T ₂₂ 、 T ₂₃ 、T ₂₄ 、T ₂₅ 、T ₂₆ 、T ₂₇ 、T ₂₈ 、T ₂₉ 、T ₃₀	トランジスタ	20
IV ₁ 、IV ₂ 、IV ₃ 、IV ₄	インバータ	
LI _n 、/LI _n 、W ₁ 、W ₂ 、LO、/LO	接続線	
M、/M	マッチ線	
SA	センスアンプ	
1、2、/2、3	タイミング信号	
TG	タイミング生成回路	
SW ₁ 、SW ₂ 、SW ₃ 、SW ₄	スイッチ	
SB ₁ 、SB ₂ 、SB ₃ 、SB ₄	スイッチボックス	
M ₁ 、M ₂ 、M ₃ 、M ₄ 、M ₅	セレクタ	30
MC ₁ 、MC ₂	メモリセルアレイ	
EN	イネーブル信号	
CLK	基準クロック信号	
ECK	偶数基準クロック	
PP、PPP	並列演算器	
CT	制御部	
MA	メモリブロック	
IS	入力セレクタ	
CL ₁ 、CL ₂ 、M ₁ 、M ₂	メモリ	40
WS	書き込みストロープ	
B	バス	
RS	読み出しストロープ	
RN	実行信号	
AD	アドレスデータ	
ST	スタート信号	
Sc ₁ 、Sc ₂ 、Sc ₃ 、Sc ₄ 、Sc ₅	切換制御信号	

【 図 1 】



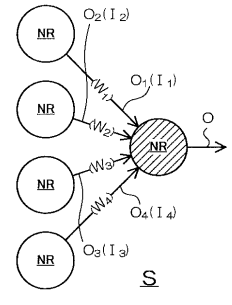
$$O = F(W_1 \times I_1 + W_2 \times I_2 + \dots + W_n \times I_n) \dots (1)$$

(a)

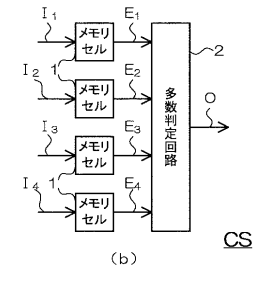


(b)

【 図 2 】



(a)

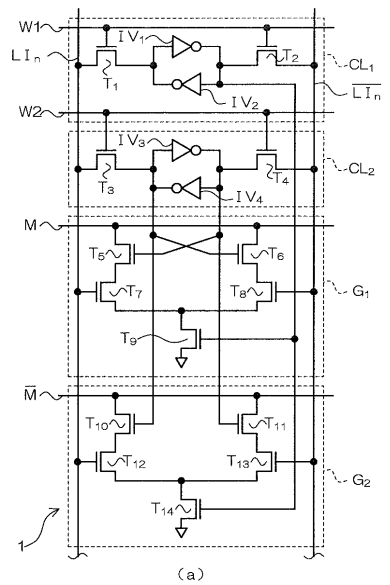


(b)

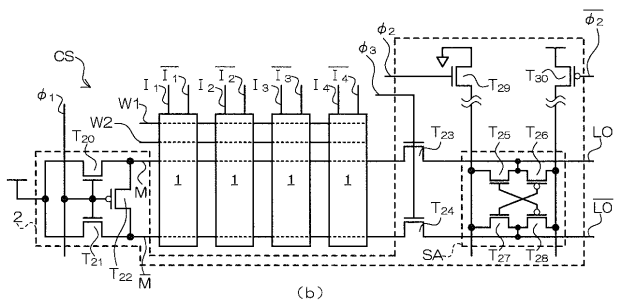
記憶値	入力データ I	出力データ E
0	0	1
0	1	0
1	0	0
1	1	1
NC	X	NC

(c)

【 図 3 】

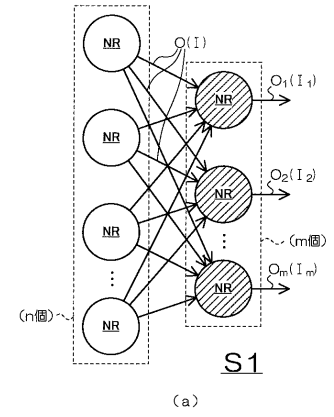


(a)

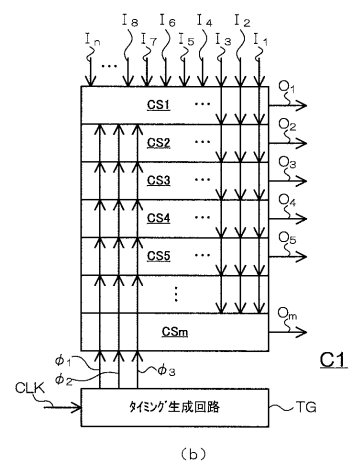


(b)

【 図 4 】

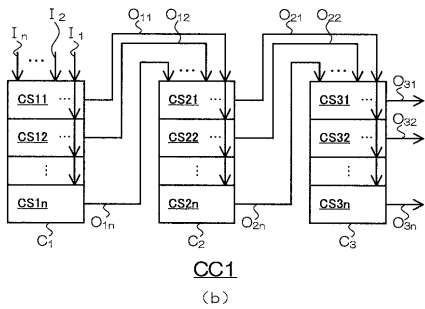
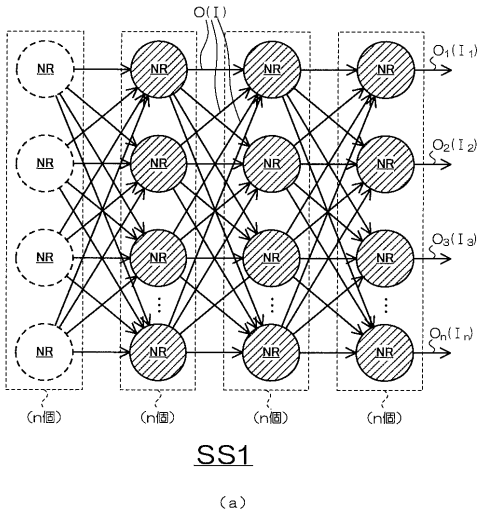


(a)

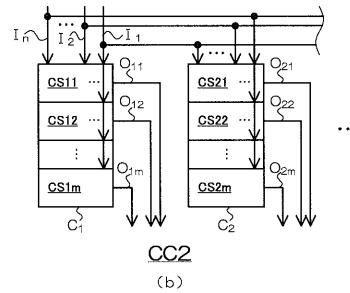
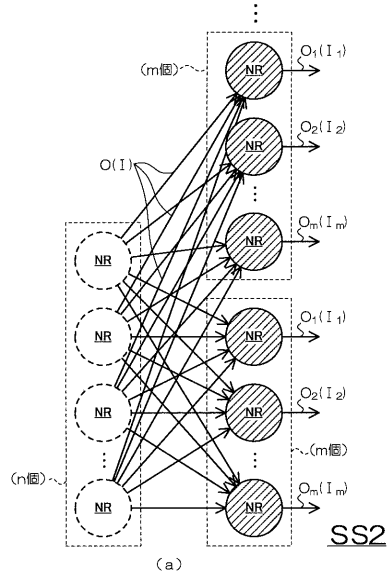


(b)

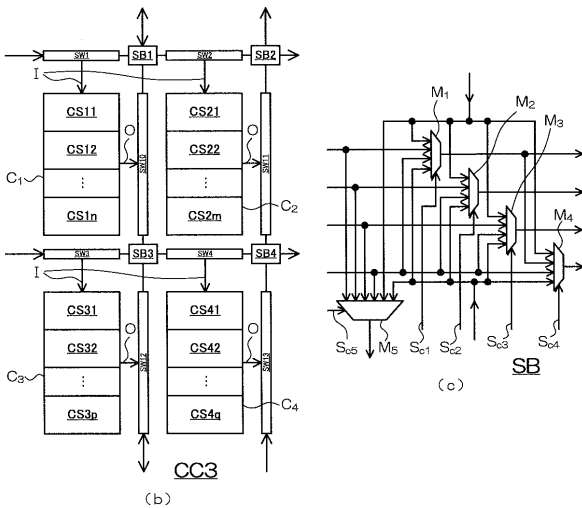
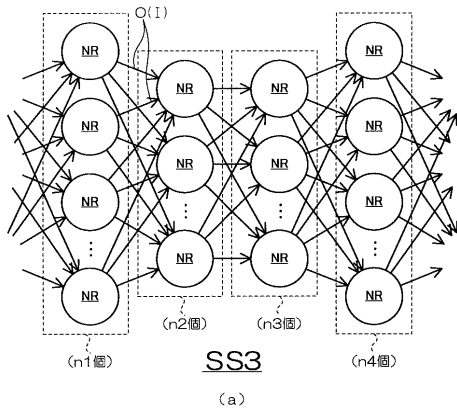
【 図 5 】



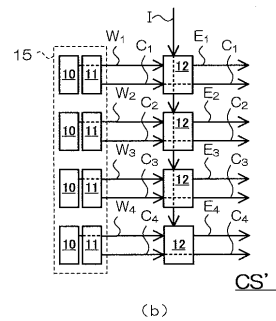
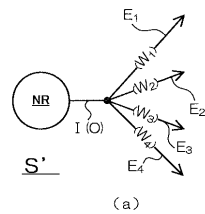
【 図 6 】



【 図 7 】



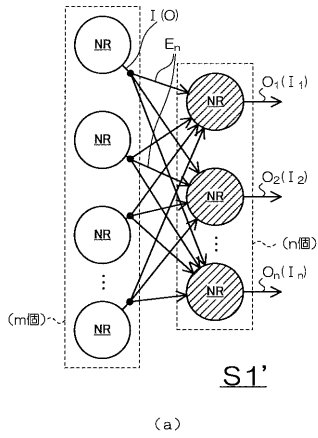
【 図 8 】



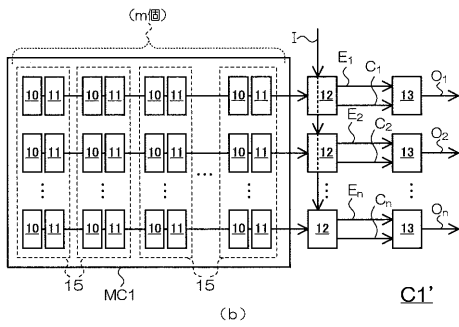
記憶値 (W)	入力値-I	XOR	XNOR 出力値(E)
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1

(c)

【図9】

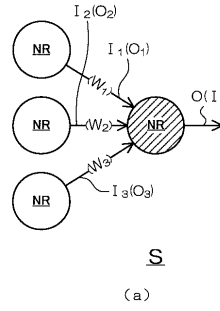


(a)

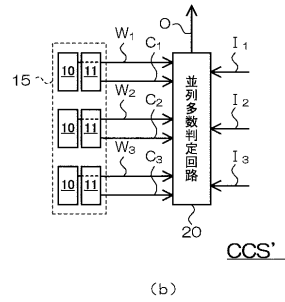


(b)

【図10】

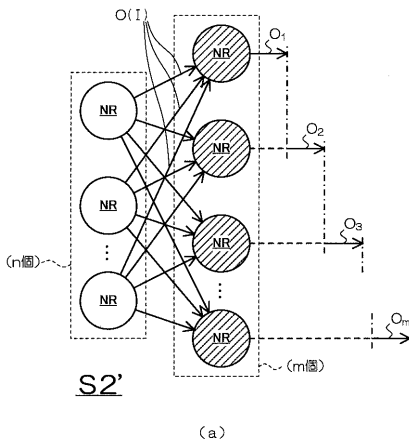


(a)

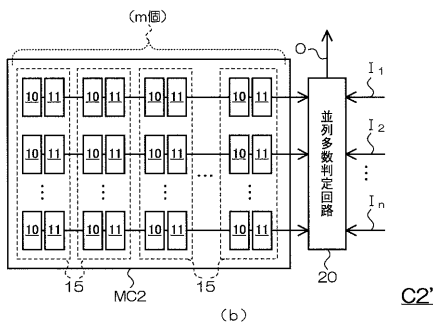


(b)

【図11】

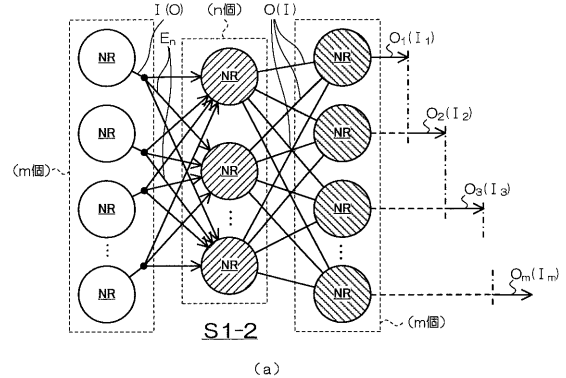


(a)

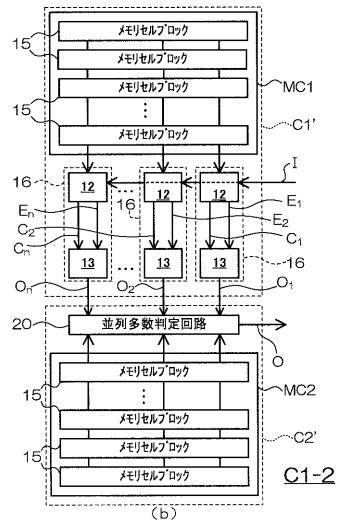


(b)

【図12】

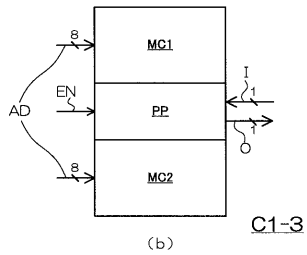
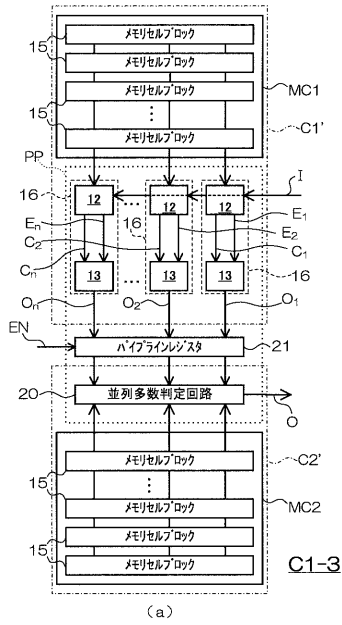


(a)

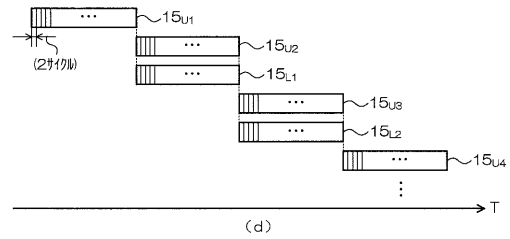
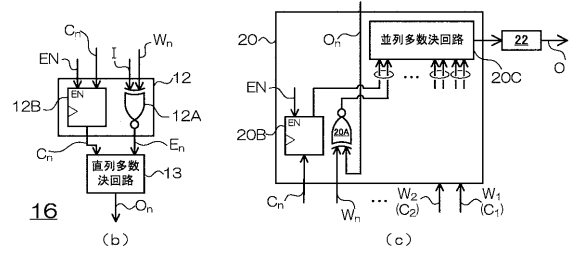
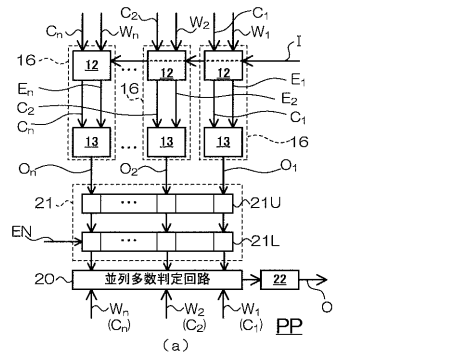


(b)

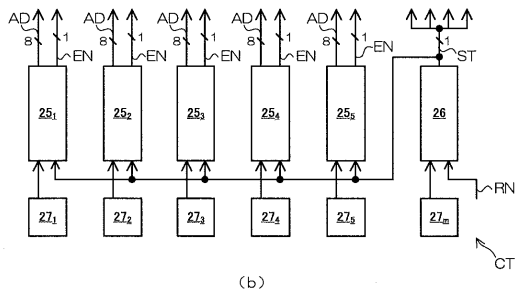
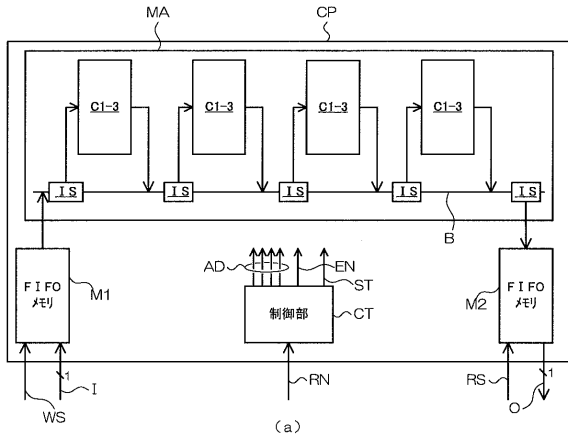
【図13】



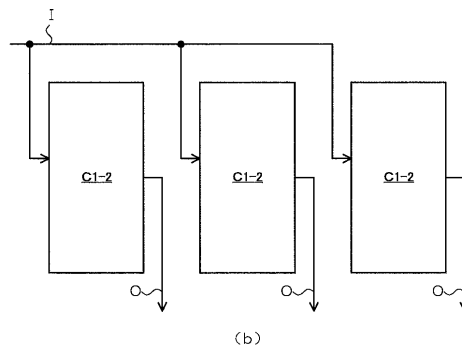
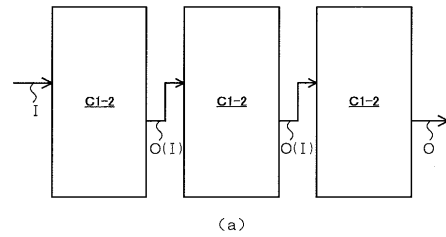
【図14】



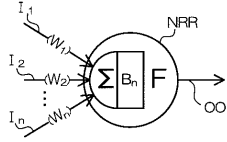
【図15】



【図16】



【図17】



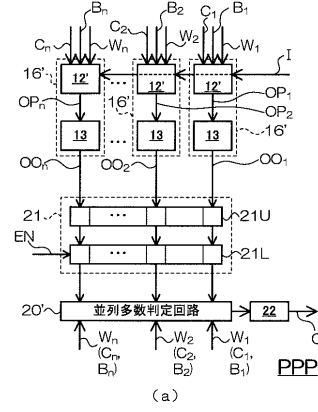
$OO = F(W_1 \times I_1 + W_2 \times I_2 + \dots + W_n \times I_n + B_n) \dots (2)$
 但し、 $0 \leq |B_n| \leq n$ 且つ B_n は整数

(a)

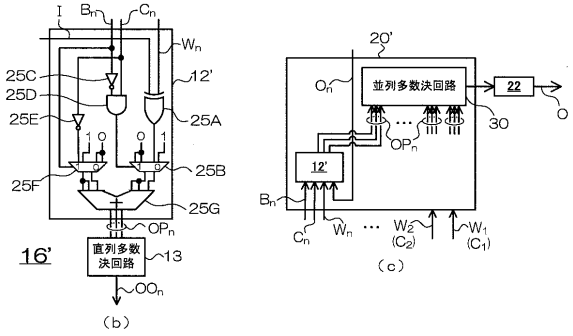
傾向係数	接続有無情報	接続	出力データ
0	0		-1 又は1
0	1		0
1	0		-2 又は0
1	1		0 又は2

(b)

【図18】



(a)



(b)

(c)