

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7292270号

(P7292270)

(45)発行日 令和5年6月16日(2023.6.16)

(24)登録日 令和5年6月8日(2023.6.8)

(51)国際特許分類

F I

G 0 6 N 3/04 (2023.01)

G 0 6 N 3/04

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00

3 5 0 C

請求項の数 13 (全27頁)

(21)出願番号 特願2020-521431(P2020-521431)  
(86)(22)出願日 平成30年10月16日(2018.10.16)  
(65)公表番号 特表2020-537787(P2020-537787  
A)  
(43)公表日 令和2年12月24日(2020.12.24)  
(86)国際出願番号 PCT/EP2018/078191  
(87)国際公開番号 WO2019/076866  
(87)国際公開日 平成31年4月25日(2019.4.25)  
審査請求日 令和3年10月14日(2021.10.14)  
(31)優先権主張番号 17290131.6  
(32)優先日 平成29年10月19日(2017.10.19)  
(33)優先権主張国・地域又は機関  
欧州特許庁(EP)

(73)特許権者 590000248  
コーニンクレッカ フィリップス エヌ  
ヴェ  
Koninklijke Philips  
N.V.  
オランダ国 5 6 5 6 アーヘー アイン  
ドーフエン ハイテック キャンパス 5 2  
High Tech Campus 52 ,  
5 6 5 6 AG Eindhoven , N  
etherlands  
(74)代理人 110001690  
弁理士法人M&Sパートナーズ  
(72)発明者  
メンコフスキ ヴラド  
オランダ国 5 6 5 6 アーヘー アイン  
ドーフエン ハイ テック キャンパス 5  
最終頁に続く

(54)【発明の名称】 ニューラルネットワークの使用

## (57)【特許請求の範囲】

## 【請求項1】

ニューロンの1つ又は複数の層から形成されたニューラルネットワークを使用するためのコンピュータ実施方法であって、

前記方法は、

前記ニューラルネットワークを使用して第1のニューラルネットワーク入力を処理し、それによって、第1のニューラルネットワーク出力を生成するステップと、

前記ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理し、それによって、第2のニューラルネットワーク出力を生成するステップと、

第1の類似性指標を決定するステップであって、第1の類似性指標が、第1のニューラルネットワーク出力と第2のニューラルネットワーク出力との間の類似性のレベルを示す、前記決定するステップと、

前記ニューラルネットワークの1つ又は複数のニューロンのセットを選択するステップであって、前記ニューロンのセットが前記ニューラルネットワークの単一層のニューロンのサブセットである、前記選択するステップと、

前記ニューラルネットワークが前記第1のニューラルネットワーク入力を処理するために使用されるとき、各選択されたニューロンの活性化値を決定するステップであって、それによって、第1の活性化値のセットを生成する、前記決定するステップと、

前記ニューラルネットワークが前記第2のニューラルネットワーク入力を処理するために使用されるとき、各選択されたニューロンの活性化値を決定するステップであって、そ

10

20

れによって、第 2 の活性化値のセットを生成する、前記決定するステップと、

前記第 1 の活性化値のセットと対応する第 2 の活性化値のセットとの間の類似性のレベルを示す第 2 の類似性指標を決定するステップと、

前記第 1 及び第 2 のニューラルネットワーク入力の中の 1 つのニューラルネットワークによる潜在的な不正確な処理を決定するステップであって、前記潜在的な不正確な処理は、

第 1 の所定の閾値を超える類似性のレベルを示す前記第 1 の類似性指標、及び第 2 の所定の閾値未満の類似性のレベルを示す前記第 2 の類似性指標、又は、

第 3 の所定の閾値未満の類似性のレベルを示す前記第 1 の類似性指標、及び第 4 の所定の閾値を超える類似性のレベルを示す前記第 2 の類似性指標  
のいずれかを含む、前記決定するステップと、

10

前記潜在的な不正確な処理をユーザにフラグで警告するステップとを有する、  
方法。

#### 【請求項 2】

前記ニューラルネットワークを使用して前記第 1 のニューラルネットワーク入力を処理するステップは、前記第 1 のニューラルネットワーク入力をある分類に分類するステップであって、その分類が第 1 のニューラルネットワーク出力である、前記分類するステップを有し、

前記ニューラルネットワークを使用して前記第 2 のニューラルネットワーク入力を処理するステップは、前記第 2 のニューラルネットワーク入力をある分類に分類するステップ  
であって、その分類が前記第 2 のニューラルネットワーク出力である、前記分類するステップを有し、

20

前記第 1 の類似性指標を決定するステップは、前記第 1 及び第 2 のニューラルネットワーク入力の分類が同じであるかどうかを識別する、前記第 1 の類似性指標を生成するステップを有し、

前記第 1 及び第 2 のニューラルネットワーク入力の中の 1 つのニューラルネットワークによる潜在的な不正確な処理を決定するステップは、

前記第 1 の類似性指標が、前記第 1 及び第 2 のニューラルネットワーク入力の分類が同じであることを示し、前記第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示しているか、又は、

30

第 1 の類似性指標が、第 1 及び第 2 のニューラルネットワーク入力の分類が異なることを示し、前記第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示しているかどうかを決定するステップを有する、

請求項 1 に記載のコンピュータ実施方法。

#### 【請求項 3】

前記ニューロンのセットは、同じ層の少なくとも 2 つのニューロンを含む、

請求項 1 又は 2 に記載のコンピュータ実施方法。

#### 【請求項 4】

前記第 2 の類似性指標を決定するステップは、第 1 の活性化値のセットと第 2 の活性化値のセットとの間の多次元距離を計算するステップを有する、

40

請求項 3 に記載のコンピュータ実施方法。

#### 【請求項 5】

前記第 2 及び第 4 の所定の閾値は、所定の閾値距離である、

請求項 1 から 4 の何れか一項に記載のコンピュータ実施方法。

#### 【請求項 6】

前記 1 つ又は複数のニューロンのセットを選択するステップは、前記第 1 のニューラルネットワーク入力を処理するために前記ニューラルネットワークが使用されるときに、活性化値の所定の範囲の外にある活性化値を有する同じ層の 1 つ又は複数のニューロンを選択するステップを有する、

請求項 1 から 5 の何れか一項に記載のコンピュータ実施方法。

50

**【請求項 7】**

前記 1 つ又は複数のニューロンのセットを選択するステップは、前記第 2 のニューラルネットワーク入力を処理するために前記ニューラルネットワークが使用されるときに、活性化値の所定の範囲の外にある活性化値を有する同じ層の 1 つ又は複数のニューロンを選択するステップを有する、

請求項 1 から 5 の何れか一項に記載のコンピュータ実施方法。

**【請求項 8】**

前記ニューラルネットワークを使用して前記第 2 のニューラルネットワーク入力を処理するステップは、前記ニューラルネットワークを使用して複数の前記第 2 のニューラルネットワーク入力を処理し、それによって、複数の第 2 のニューラルネットワーク出力を生成するステップを有し、

10

前記第 1 の類似性指標を決定するステップは、前記複数の第 2 のニューラルネットワーク入力の各々に対して、前記第 1 のニューラルネットワーク出力と前記第 2 のニューラルネットワーク出力との間の類似性のレベルを示す第 1 の類似性指標を決定し、それによって、複数の第 1 の類似性指標を生成するステップを有し、

前記第 2 の活性化値のセットを生成するステップは、前記複数の第 2 のニューラルネットワーク入力の各々に対して、前記第 2 のニューラルネットワーク入力を処理するために前記ニューラルネットワークが使用されるときに各選択されたニューロンの活性化値を決定し、それによって、前記第 2 のニューラルネットワーク入力ごとに第 2 の活性化値のセットを生成し、それによって、前記複数の第 2 の活性化値のセットを生成するステップを有し、

20

前記第 2 の類似性指標を決定するステップは、前記複数の第 2 のニューラルネットワーク入力の各々に対して、前記第 1 の活性化値のセットと対応する前記第 2 の活性化値のセットとの間の類似性のレベルを示す前記第 2 の類似性指標を決定し、それによって、複数の第 2 の類似性指標を生成するステップを有し、

前記第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つのニューラルネットワークによる潜在的な不正確な処理を決定するステップが、

前記第 2 のニューラルネットワーク入力ごとに、

前記第 1 の類似性指標が第 1 の所定の閾値を超える類似性のレベルを示し、前記第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示すか、又は、

30

前記第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示し、前記第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示すかどうかを決定し、

それによって、前記第 2 のニューラルネットワーク入力ごとに、前記第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つが前記ニューラルネットワークによって潜在的に不正確に処理されたかどうかを決定するステップと、

前記第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つが前記ニューラルネットワークによって潜在的に不正確に処理されたことが決定されるすべての前記第 2 のニューラルネットワーク入力を識別するステップとを有する、

請求項 1 又は 2 に記載のコンピュータ実施方法。

**【請求項 9】**

40

前記第 2 のニューラルネットワーク入力ごとに第 2 の類似性指標を決定するステップは、

前記第 1 の活性化値のセットと前記複数の第 2 の活性化値のセットの各々とに多次元クラスタリングを実行するステップと、

前記第 2 のニューラルネットワーク入力ごとに、前記第 1 のニューラルネットワーク入力及び第 2 のニューラルネットワーク入力と同じクラスタの活性化値のセットに関連しているかどうかを示す第 2 の類似性指標を生成するステップとを有する、

請求項 8 に記載の方法。

**【請求項 10】**

前記第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つ又は複数に対して正しいニューラルネットワーク出力を示すユーザ入力を得て、前記正しいニューラルネットワー

50

ク出力に基づいて前記ニューラルネットワーク入力を再訓練するステップをさらに有する、請求項 1 から 9 の何れか一項に記載のコンピュータ実施方法。

【請求項 1 1】

前記第 1 及び第 2 のニューラルネットワーク入力は、医用画像、医療機器若しくはセンサからの出力、又は医療記録のうちの任意の 1 つ又はこれらの任意の組合せを含む、請求項 1 から 10 の何れか一項に記載のコンピュータ実施方法。

【請求項 1 2】

非一時的コンピュータ可読媒体の中に具現されたコンピュータ可読コードを有し、前記コンピュータ可読コードは、適切なコンピュータ又はプロセッサによる実行により、前記コンピュータ又は前記プロセッサに請求項 1 から 10 の何れか一項に記載の方法を実行させるように構成される、非一時的コンピュータ可読媒体。

10

【請求項 1 3】

命令のセットを表す命令データを含むメモリと、前記メモリと通信し、前記命令のセットを実行するプロセッサとを含む、システムであって、

前記命令のセットは、前記プロセッサによって実行されると、

ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理して、第 1 のニューラルネットワーク出力を生成すること、

前記ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理して、第 2 のニューラルネットワーク出力を生成すること、

第 1 の類似性指標を決定することであって、前記第 1 の類似性指標が、前記第 1 のニューラルネットワーク出力と前記第 2 のニューラルネットワーク出力との間の類似性のレベルを示す、決定すること、

20

前記ニューラルネットワークの 1 つ又は複数のニューロンのセットを選択することであって、前記ニューロンのセットが前記ニューラルネットワークの単一層のニューロンのサブセットである、選択すること、

前記第 1 のニューラルネットワーク入力を処理するために前記ニューラルネットワークが使用されるときに、各選択されたニューロンの活性化値を決定し、それによって、第 1 の活性化値のセットを生成すること、

前記第 2 のニューラルネットワーク入力を処理するために前記ニューラルネットワークが使用されるときに、各選択されたニューロンの活性化値を決定し、それによって、第 2 の活性化値のセットを生成すること、

30

前記第 1 の活性化値のセットと対応する第 2 の活性化値のセットとの間の類似性のレベルを示す第 2 の類似性指標を決定すること、

前記第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つのニューラルネットワークによる潜在的な不正確な処理を決定することであって、前記潜在的な不正確な処理は、

前記第 1 の類似性指標が第 1 の所定の閾値を超える類似性のレベルを示し、前記第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示すか、又は

前記第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示し、第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示すことを含む、決定すること、並びに、

40

前記潜在的な不正確な処理をユーザにフラグで警告することを、プロセッサに行わせる、システム。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書に記載される様々な実施形態は、ニューラルネットワークの使用の分野に関する。より詳細には、しかし限定ではなく、様々な実施形態は、1 つ又は複数の層から形成されたニューラルネットワークを使用するシステム及び方法に関する。

【背景技術】

【0002】

50

新しい技術を採用することになる場合、特に、例えば、その技術を使用してヘルスケア意思決定を行うヘルスケアシナリオでは、ユーザが新しい技術を信頼することが重要である。ユーザは、従来のシステムを信頼する傾向があるが、その理由は、システムの動作が全く透明であり、それゆえに、システムの挙動が合理的に予測可能であるからである。

【発明の概要】

【発明が解決しようとする課題】

【0003】

機械学習では、多くの場合、確度と了解度との間にトレードオフがある。ディープニューラルネットワークは、特に、データの有用な意味表現の学習に成功することが実証されている。しかしながら、ディープニューラルネットワークは、複雑なアーキテクチャ及びプロセスを有しており、それは、ユーザがニューラルネットワークで実行される意思決定プロセスを理解するのを困難にしている。それゆえに、ディープラーニングベースのシステムは、複雑な構造及びデータ駆動型設計プロセスのために透明性及び予測性を欠く傾向がある。

10

【0004】

入力データの処理におけるこの透明性の欠如の結果として、ユーザは、プロセスの背後の出力又は方法の妥当性を簡単に検証することができないので、ニューラルネットワークの出力を信用しなくなることがある。さらに、ユーザは出力に異議があることがあり、ニューラルネットワークがどのように意思決定に至ったかを知ることなしには、これにより、システムへの不信を生じることがある。その結果、これは、実際に非常に正確な高い性能の技術であるものについて使用の停止又は採用の縮小をもたらすことがある。

20

【0005】

それゆえに、上記の問題点を改善する方法及びシステムが必要である。

【0006】

ニューラルネットワークは、層状に配列された人工ニューロン又は単に「ニューロン」と呼ばれる多数のユニットから形成され、多数のユニットは、脳のニューロンをシミュレートするように構成される。ニューラルネットワークモデルの各層は、入力に様々な変換を実行する。ニューラルネットワークのニューロンは、1つ又は複数の接続又はエッジによってネットワークの1つ又は複数の他のニューロンに接続される。各ニューロン又はエッジは、関連する重み（バイアスと呼ばれることもある）を有し、関連する重みは、ニューロンからの出力に、又は1つのニューロンから接続済みニューロンまでの接続に適用される。重みは、例えば訓練データを使用して、ニューラルネットワークの訓練により更新される。信号（例えば、データ）は、ニューラルネットワークの最初の入力層からネットワークの最後の出力層までの進み、1つ又は複数の中間の「隠れた」層を横切る。訓練により、重みは、ニューラルネットワークの出力が期待出力に近づくように更新される。

30

【課題を解決するための手段】

【0007】

第1の態様によれば、ニューロンの1つ又は複数の層から形成されたニューラルネットワークを使用するためのコンピュータ実施方法がもたらされ、この方法は、ニューラルネットワークを使用して第1のニューラルネットワーク入力を処理し、それによって、第1のニューラルネットワーク出力を生成するステップと、ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理し、それによって、第2のニューラルネットワーク出力を生成するステップと、第1の類似性指標を決定するステップであり、第1の類似性指標が、第1のニューラルネットワーク出力と第2のニューラルネットワーク出力との間の類似性のレベルを示す、決定するステップとを有する。この方法は、ニューラルネットワークの1つ又は複数のニューロンのセットを選択するステップであって、ニューロンのセットがニューラルネットワークの単一層のニューロンのサブセットである、選択するステップと、ニューラルネットワークが第1のニューラルネットワーク入力を処理するために使用されるとき、各選択されたニューロンの活性化値を決定するステップであって、それによって、第1の活性化値のセットを生成する、前記決定するステップと、ニュー

40

50

ーラルネットワークが第2のニューラルネットワーク入力を処理するために使用されるとき、各選択されたニューロンの活性化値を決定するステップであって、それによって、第2の活性化値のセットを生成する、前記決定するステップと、第1の活性化値のセットと対応する第2の活性化値のセットとの間の類似性のレベルを示す第2の類似性指標を決定するステップとをさらに有する。この方法は、第1の類似性指標が第1の所定の閾値を超える類似性のレベルを示し、第2の類似性指標が第2の所定の閾値未満の類似性のレベルを示すか、又は第1の類似性指標が第3の所定の閾値未満の類似性のレベルを示し、第2の類似性指標が第4の所定の閾値を超える類似性のレベルを示すかどうかを決定し、それによって、第1及び第2のニューラルネットワーク入力のうちの1つがニューラルネットワークによって不正確に処理されたかどうかを予測するステップをさらに有する。

10

【0008】

したがって、実施形態は、有利には、第1又は第2のニューラルネットワーク入力のうちの1つが不正確に処理されたかどうかを認識又は予測する。言い換えれば、ニューラルネットワークが不正確なニューラルネットワーク出力を生成したかどうかに関して予測が行われる。不正確なニューラルネットワーク出力は、ニューラルネットワーク入力に基づくニューラルネットワークの意図した又は所望の出力を反映していないもの（例えば、ニューラルネットワーク入力の誤分類、又は問題への不十分な最適解決策）である。

【0009】

さらに言い換えると、実施形態は、2つのニューラルネットワーク入力の処理間の不整合を識別できるようにし、それによって、ニューラルネットワークの確度又はエラーに関する評価を実行できるようにする。したがって、ニューラルネットワークが、ニューラルネットワーク入力に所望のタスクを正確に実行しているかどうか、すなわち、正しい又は正確なニューラルネットワーク出力を生成しているかどうかに関して予測を行うことができる。

20

【0010】

これは、有利には、ニューラルネットワークの再訓練を促すために又は不正確なニューラルネットワーク出力の可能性にユーザの注意を引きつけるために（例えば、ユーザがニューラルネットワークを正確であると信じないようにするために）使用することができる。

【0011】

類似性指標は、2つの値又は値のセット間の類似性のレベルを示す（離散的、絶対的、又は連続的）尺度である。例として、2値指標は、2つの値（のセット）が類似しているか又は類似していないと見なされるかどうかを示す。別の例では、2つの値（のセット）間の距離が類似性指標を形成し、（一般に）距離が小さいほど、類似性が大きい。

30

【0012】

実施形態は、ニューラルネットワークのニューロンのサブセットのみを使用して、2つのニューラルネットワーク入力間の比較のための活性化値のセットを生成することにより、層に含まれる具体的な特定の情報を比較することができるという理解に依拠する。例えば、層中のニューロンの第1のサブセットは、ニューラルネットワークの特定のタスクに関連し、一方、第2のサブセットは異なるタスクに関連する。したがって、特定のタスクがニューラルネットワーク入力の処理の際に不正確を引き起こしているかどうかに関して識別を行い、それによって、ニューラルネットワークの限界及び/又は基礎をなすプロセスについての理解の改善を可能にし、又はニューラルネットワークの管理された再訓練を可能にする。特に、ニューラルネットワークの決定プロセスについてのより完全な知識がユーザに提供される。

40

【0013】

実施形態では、ニューラルネットワークを使用して第1のニューラルネットワーク入力を処理するステップは、第1のニューラルネットワーク入力のある分類に分類するステップであって、その分類が第1のニューラルネットワーク出力である、分類するステップを有し、ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理するステップは、第2のニューラルネットワーク入力のある分類に分類するステップであって

50

、その分類が第2のニューラルネットワーク出力である、分類するステップを有し、第1の類似性指標を決定するステップは、第1及び第2のニューラルネットワーク入力の分類が同じであるかどうかを識別する第1の類似性指標を生成するステップを有し、第1及び第2のニューラルネットワーク入力のうちの1つが不正確に処理されたかどうかを予測するステップは、第1の類似性指標が、第1及び第2のニューラルネットワーク入力の分類が同じであることを示し、第2の類似性指標が第2の所定の閾値未満の類似性のレベルを示しているか、又は第1の類似性指標が、第1及び第2のニューラルネットワーク入力の分類が異なることを示し、第2の類似性指標が第4の所定の閾値を超える類似性のレベルを示しているかどうかを決定し、それによって、第1及び第2のニューラルネットワーク入力のうちの1つがニューラルネットワークによって不正確に分類されたかどうかを予測するステップを有する。

10

【0014】

言い換えれば、ニューラルネットワークはニューラルネットワーク入力の分類を実行するように構成され、分類が類似しているかどうかの決定は分類が同一であるかどうかの決定を含む。したがって、ニューラルネットワークがニューラルネットワーク入力のうちの1つを不正確に処理したかどうかの決定又は予測は、ニューラルネットワークがニューラルネットワーク入力のうちの1つを誤って分類したかどうかの予測を含む。

【0015】

好ましくは、ニューロンのセットは、同じ層の少なくとも2つのニューロンを含む。

【0016】

第2の類似性指標を決定するステップは、第1の値のセットと第2の値のセットとの間の多次元距離を計算するステップを有する。これは、2つの値のセット間の類似性を示す値を決定する信頼できる方法を提供する（多次元距離のサイズが類似性を表すので）。それゆえに、そのような実施形態では、第2及び第4の所定の閾値は、所定の閾値距離である。

20

【0017】

1つ又は複数のニューロンのセットを選択するステップは、好ましくは、第1のニューラルネットワーク入力を処理するためにニューラルネットワークが使用されるときに、活性化値の所定の範囲の外にある活性化値を有する同じ層の1つ又は複数のニューロンを選択するステップを有する。代替として、1つ又は複数のニューロンのセットを選択するステップは、第2のニューラルネットワーク入力を処理するためにニューラルネットワークが使用されるときに、活性化値の所定の範囲の外にある活性化値を有する同じ層の1つ又は複数のニューロンを選択するステップを有する。

30

【0018】

したがって、異常な活性化値は、比較のために選択されるニューロンのセットを定義する。これにより、層の潜在的に重要なニューロンを識別する可能性が高まる。

【0019】

好ましくは、ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理するステップは、ニューラルネットワークを使用して複数の第2のニューラルネットワーク入力を処理し、それによって、複数の第2のニューラルネットワーク出力を生成するステップを有し、第1の類似性指標を決定するステップは、複数の第2のニューラルネットワーク入力の各々に対して、第1のニューラルネットワーク出力と第2のニューラルネットワーク出力との間の類似性のレベルを示す第1の類似性指標を決定し、それによって、複数の第1の類似性指標を生成するステップを有し、第2の活性化値のセットを生成するステップは、複数の第2のニューラルネットワーク入力の各々に対して、第2のニューラルネットワーク入力を処理するためにニューラルネットワークが使用されるときに各選択されたニューロンの活性化値を決定し、それによって、第2のニューラルネットワーク入力ごとに第2の活性化値のセットを生成し、それによって、複数の第2の活性化値のセットを生成するステップを有し、第2の類似性指標を決定するステップは、複数の第2のニューラルネットワーク入力の各々に対して、第1の活性化値のセットと対応する第2

40

50

の活性化値のセットとの間の類似性のレベルを示す第2の類似性指標を決定し、それによって、複数の第2の類似性指標を生成するステップを有し、第1及び第2のニューラルネットワーク入力の中の1つがニューラルネットワークによって不正確に処理されたかどうかを予測するステップが、第2のニューラルネットワーク入力ごとに、第1の類似性指標が第1の所定の閾値を超える類似性のレベルを示し、第2の類似性指標が第2の所定の閾値未満の類似性のレベルを示すか、又は第1の類似性指標が第3の所定の閾値未満の類似性のレベルを示し、第2の類似性指標が第4の所定の閾値を超える類似性のレベルを示すかどうかを決定し、それによって、第2のニューラルネットワーク入力ごとに、第1及び第2のニューラルネットワーク入力の中の1つがニューラルネットワークによって不正確に処理されたかどうかを予測するステップと、第1及び第2のニューラルネットワーク入力の中の1つがニューラルネットワークによって不正確に処理されたことが予測されるすべての第2のニューラルネットワーク入力を識別するステップとを有する。

10

【0020】

言い換えれば、1つを超える第2のニューラルネットワーク入力は、第1のニューラルネットワークに対して少なくとも1つの差異（例えば、類似していないニューラルネットワーク出力又は類似していない活性化値のセットのいずれか）を有する第2のニューラルネットワーク入力の集合又はグループを識別するために、第1のニューラルネットワーク入力に対して評価される。これにより、潜在的に不正確に処理されたニューラルネットワーク入力の大きいセットを識別することができる。

20

【0021】

当然、これが、不正確に処理された第1のニューラルネットワーク入力であることが理解されよう。そのような決定は、多数又は過半数の第2のニューラルネットワーク入力（複数のニューラルネットワーク入力の中の）が、潜在的に不正確に処理されているとして識別される場合に行われる。

【0022】

第2のニューラルネットワーク入力ごとに第2の類似性指標を決定するステップは、第1の活性化値のセットと複数の第2の活性化値のセットの各々とに多次元クラスタリングを実行するステップと、第2のニューラルネットワーク入力ごとに、第1のニューラルネットワーク入力及び第2のニューラルネットワーク入力と同じクラスタの活性化値のセットに関連しているかどうかを示す第2の類似性指標を生成するステップとを有する。

30

【0023】

したがって、第2の類似性指標は、第1の活性化値のセットが第2の活性化値のセットと同じクラスタにあるかどうかを示す。クラスタリングを実行するために、総数の多い第2の活性化値のセット、それゆえに、第2のニューラルネットワーク入力（例えば、50個を超える又は100個を超える）が必要とされる。

【0024】

いくつかの実施形態では、この方法は、第1及び第2のニューラルネットワーク入力の中の1つがニューラルネットワークによって不正確に処理されたことを予測したことに応じて、ニューラルネットワークで訓練プロセスを実行するステップをさらに有する。

【0025】

40

したがって、第1及び第2のニューラルネットワーク入力の中の1つが不正確に処理されたことの識別が、訓練プロセスを実行するように促す。訓練プロセスは、例えば、ユーザ入力（例えば、第1の/第2のニューラルネットワーク入力に対して正しいニューラルネットワーク出力を示す）に基づくか、又はニューラルネットワークのための訓練データにさらに基づく。

【0026】

他の例では、潜在的に不正確に処理されたニューラルネットワーク入力（及び/又は関連するニューラルネットワーク出力）は、ユーザによるレビューのためにフラグで警告される。これは、例えば、アラートの生成を含む。

【0027】

50



第2の態様によれば、非一時的コンピュータ可読媒体を含むコンピュータプログラム製品がもたらされ、コンピュータ可読媒体は、その中に具現されたコンピュータ可読コードを有し、コンピュータ可読コードは、適切なコンピュータ又はプロセッサによる実行により、コンピュータ又はプロセッサに本明細書に記載された実施形態のうちのいずれかについての方法を実行させるように構成される。

【0028】

第3の態様によれば、命令のセットを表す命令データを含むメモリと、メモリと通信し命令のセットを実行するように構成されたプロセッサとを含むシステムがある。命令のセットは、プロセッサによって実行されると、ニューラルネットワークを使用して第1のニューラルネットワーク入力を処理し、それによって、第1のニューラルネットワーク出力を生成すること、ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理し、それによって、第2のニューラルネットワーク出力を生成すること、第1の類似性指標を決定することであって、第1の類似性指標が、第1のニューラルネットワーク出力と第2のニューラルネットワーク出力との間の類似性のレベルを示す、決定すること、ニューラルネットワークの1つ又は複数のニューロンのセットを選択することであって、ニューロンのセットがニューラルネットワークの単一層のニューロンのサブセットである、選択すること、ニューラルネットワークを使用して第1のニューラルネットワーク入力を処理し、それによって、第1の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定すること、ニューラルネットワークを使用して第2のニューラルネットワーク入力を処理し、それによって、第2の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定すること、第1の活性化値のセットと対応する第2の活性化値のセットとの間の類似性のレベルを示す第2の類似性指標を決定すること、第1の類似性指標が第1の所定の閾値を超える類似性のレベルを示し、第2の類似性指標が第2の所定の閾値未満の類似性のレベルを示すか、又は第1の類似性指標が第3の所定の閾値未満の類似性のレベルを示し、第2の類似性指標が第4の所定の閾値を超える類似性のレベルを示すかどうかを決定し、それによって、第1及び第2のニューラルネットワーク入力のうちの1つがニューラルネットワークによって不正確に処理されたかどうかを予測することをプロセッサに行わせる。

【0029】

実施形態のよりよい理解のために、及び実施形態がどのように実行に移されるかをより明確に示すために、次に、単に例として、添付の図面を参照する。

【図面の簡単な説明】

【0030】

【図1】一実施形態の文脈理解のためのニューラルネットワークを示す図である。

【図2】一実施形態の文脈理解のためのニューラルネットワークを示す図である。

【図3】一実施形態によるコンピュータ実施方法を示す図である。

【図4】ユーザに表示されるような、一実施形態による方法の例示の出力を示す図である。

【図5】一実施形態によるシステムを示す図である。

【発明を実施するための形態】

【0031】

既存の問題のうちのいくつかを克服する、ニューラルネットワークへの入力データを処理するための改善された方法及びシステムが提供される。

【0032】

本発明の概念によれば、ニューラルネットワークによって誤って処理されたニューラルネットワークのニューラルネットワーク入力を識別するための方法、システム、及びコンピュータプログラム製品が提案される。ニューラルネットワーク入力に関連する活性化値のセット（単一層のニューロンのサブセットの）が得られる。ニューラルネットワーク入力に関連するニューラルネットワーク出力も得られる。第1及び第2のニューラルネットワーク入力が、類似している活性化値のセットを共有するが、類似していないニューラルネットワーク出力を共有するかどうか、又はその逆であるかに関して決定が行われる。こ

のようにして、第 1 及び第 2 のニューラルネットワーク入力の中の 1 つが、ニューラルネットワークによって誤って処理されたかどうかに関して予測が行われる。

【 0 0 3 3 】

実施形態は、類似しているニューラルネットワーク出力に関連するニューラルネットワーク入力と同様のやり方で処理されるべきであるという認識に少なくとも部分的に基づく。特に、単一層のニューロンのサブセットのみを使用することによって、ニューラルネットワークによって実行されるプロセスの特定のプロセス段階が、確度に関して評価される。

【 0 0 3 4 】

例示の実施形態は、例えば、追加の訓練を必要とするニューラルネットワークを識別するために、又はニューラルネットワーク出力の確度を予測するために利用される。そのような情報は、医療従事者には特に重要であり、ニューラルネットワークによる分類などの出力を使用して患者の症状を診断又は評価する（例えば、肺 CT スキャンなどの画像データを処理することによって）際に医師の臨床医を支援することができる。

【 0 0 3 5 】

想定されているニューラルネットワークは、ニューラルネットワーク入力を処理して出力を提供するための任意のタイプのニューラルネットワーク、例えば、ニューラルネットワーク入力を分類するために使用されるニューラルネットワークである。

【 0 0 3 6 】

ニューラルネットワークは、人工ニューラルネットワーク、特にディープニューラルネットワークとすることができる。人工ニューラルネットワーク又は単にニューラルネットワークは、当業者によく知られているが、手短に言えば、ニューラルネットワークは、出力データを生成するために入力データを処理する、例えば、（医療）画像などの入力データを分類するために使用される 1 つのタイプのモデルである。

【 0 0 3 7 】

いくつかの実施形態では、ニューラルネットワークは、ニューラルネットワーク入力を 1 つ又は複数のタイプのデータとして分類する（例えば、ニューラルネットワーク入力を医療記録などの特定のタイプの医療データとして分類する）際に使用するためのニューラルネットワークである。いくつかの実施形態では、ニューラルネットワークは、ニューラルネットワーク入力の内容に応じてニューラルネットワーク入力を分類する。例えば、ニューラルネットワーク入力画像である実施形態では、ニューラルネットワークは、画像に示されるもの（例えば、「心臓」又は「肺」）に応じて画像を分類する。いくつかの実施形態では、ニューラルネットワークは、ニューラルネットワーク入力の内容の解釈に応じてニューラルネットワーク入力を分類する（例えば、画像中の構造が、「癌性」又は「非癌性」として分類される）。

【 0 0 3 8 】

図 1 は、明瞭さ及び文脈のために、例示のニューラルネットワーク 1 の表現を示す。

【 0 0 3 9 】

一般的に言えば、ニューラルネットワークの構造は、人間の脳によって着想を与えられている。ニューラルネットワーク 1 は層 2、3、及び 4 から構成され、各層は、複数のニューロン  $2a \sim 2d$ 、 $3a \sim 3d$ 、 $4a \sim 4d$  を含む。各ニューロンは、数学的演算を含む。ニューラルネットワーク入力  $d_i$ （例えば、入力ノード  $n_i$  で受け取られた）を処理するプロセスでは、各ニューロンの数学的演算がニューラルネットワーク入力に実行されて、数値出力が生成され、ニューラルネットワークの各層の出力は、次の層に順次供給される。次いで、最終層 4 は、例えば、出力ノード  $n_o$  においてニューラルネットワーク出力  $d_o$  を供給する（例えば、プーリングステップにおいて）。ニューロンの数値出力の大きさ（ニューラルネットワーク入力を分類する場合）は、多くの場合、そのニューロンの「活性化値」又は「活性化レベル」と呼ばれる。したがって、活性化値又は活性化レベルは、ニューラルネットワーク入力の処理中のニューロンの活性化の量を表す数値である。

【 0 0 4 0 】

ニューラルネットワークの異なる層は、当技術分野でよく知られているように、異なる

10

20

30

40

50

機能を実行する。例えば、所与の層は、畳み込み層、プーリング層、正規化線形ユニット、全結合層、部分的結合層、又は損失層である。異なる層の他の可能な機能が、当業者にはよく知られているであろう。

【0041】

畳み込みニューラルネットワークなどのいくつかのニューラルネットワークでは、ニューラルネットワークの下位層（すなわち、ニューラルネットワークの一連の層の先頭に近い層）は、ニューラルネットワーク入力が分類されるときに小さい特徴又はパターン（すなわち、それに応じた出力）によって活性化され、一方、上位層（すなわち、ニューラルネットワークの一連の層の終わりに近い層）は、ニューラルネットワーク入力処理されるときに漸増的に大きい特徴によって活性化される。

10

【0042】

一例として、ニューラルネットワーク入力  $d_i$  が画像を含み、ニューラルネットワークが画像を分類するように構成されている場合、ニューラルネットワークの下位層は、小さい特徴（例えば、画像のエッジパターンなど）によって活性化され、中間レベル層は、例えばより大きい形状及び形態などの画像の特徴によって活性化され、一方、出力に最も近い層（例えば、上位層）は、画像の対象全体によって活性化される。異なるニューラルネットワーク入力は、異なる活性化パターンを作り出す（例えば、ネットワーク内に異なる活性化シグネチャを有する）。例えば、ニューラルネットワークが画像の分類を実行するように構成されている場合、心臓の画像は、肺の画像とは異なる活性化パターンを生成する。したがって、ニューラルネットワークは、ニューラルネットワークで生成された活性化パターンに応じてニューラルネットワーク入力を処理する。

20

【0043】

一般に、ニューロンは、数学的演算がそのニューロンの正常な範囲（例えば、統計的に平均又は近似的に平均）の外の出力を生成する場合、「高い」又は「強い」活性化を有すると見なされる。逆に、ニューロンは、数学的演算が、そのニューロンの正常な範囲（例えば、統計的に平均又は近似的に平均）に含まれる出力を生成する場合、「低い」又は「弱い」活性化を有すると見なされる。正常範囲は統計的に定義される（例えば、ニューロンがニューラルネットワーク入力を分類する場合、ニューロンの平均活性化値からの特定の数の標準偏差）。ニューロンの活性化値を比較する別のやり方は、ニューロンの活性化値をニューラルネットワークの他のニューロンと比較することである。例えば、ニューロンの活性化は、同じ層の他のニューロンの活性化の大きさと比べて高い又は低いと見なされる。いくつかの実施形態では、ニューロンは、ニューロンの活性化値（例えば、数値出力）が同じ層の他のニューロンの活性化値と比較して統計的に範囲外にある場合、高度に活性化されていると見なされる。統計的な外れ値は、当業者には理解されるような多くの異なるやり方で定義される。一例として、ニューロンは、活性化値が、同じ層のニューロンの平均活性化値から数  $\times$  の標準偏差よりも大きい場合、統計的に範囲外にある活性化値を有する。当業者は、統計的な外れ値を決定する他のやり方をよく知っているであろう。

30

【0044】

いくつかの例では、ニューロンの活性化の範囲（例えば、ニューロンによって出力される数値範囲）は、そのニューロンによって実行される数学的演算のタイプに依存する。あるニューロン（例えば、シグモイド関数を利用するニューロン）は、0と1との間で活性化され（例えば、出力を生成し）、一方、他のものは、無限の活性化値を有する（例えば、 $-$  と  $+$  との間のどこかで出力を生成する）。

40

【0045】

本明細書のいくつかの例では、ニューラルネットワーク入力  $d_i$  が画像を含み、ニューラルネットワークが画像の内容を分類するためのものである場合、ニューラルネットワークの各ニューロンは、画像の画素（又は3次元ではボクセル）値の重み付き線形和とそれに続く非線形変換を含む数学的演算を含む。ニューラルネットワークで使用される非線形変換の例には、シグモイド関数、双曲線正接関数、及び正規化線形関数が含まれる。ニューラルネットワークの各層のニューロンは、一般に、単一のタイプの変換の異なる重み付

50

き組合せ（例えば、同じタイプの変換、シグモイドなどであるが、異なる重み付け）を含む。当業者にはよく知られているように、いくつかの層では、同じ重みが、線形和において各ニューロンによって適用され、これは、例えば、畳み込み層の場合に該当する。各ニューロンの出力は数値であり、ニューロンの数値出力の大きさは、画像を分類するために使用されるニューロン活性化パターンを形成する。

【0046】

本発明は、ニューラルネットワーク出力  $d_o$  を生成するためのニューラルネットワーク入力  $d_i$  の処理が不正確に実行されたかどうかの識別を支援するための概念を提案する。言い換えれば、この概念は、ニューラルネットワーク出力が誤っているかどうかを予測することを提案する。この概念は、以下で説明するように、コンピュータ実施方法、システム、又はコンピュータプログラム製品によって実現される。

10

【0047】

発明者等は、類似しているニューラルネットワーク出力  $d_o$  に関連するニューラルネットワーク入力  $d_i$  はニューラルネットワークによって同様のやり方で処理されるべきであることを認識した。したがって、異なるニューラルネットワーク入力に関連するニューラルネットワーク出力間に十分な類似性がある場合、異なるニューラルネットワーク入力の処理の間、ニューロンの活性化値間にも類似性があるべきであることが認識された。非類似性は、ニューラルネットワーク入力のうちの1つの不正確な処理を示す。

【0048】

発明者等は、単一層のニューロンのサブセットのみの選択により（異なるニューラルネットワーク入力間のニューロンの活性化値を比較するために）、あり得る不正確がどこで生じたかに関する追加の情報と、それによって、ニューラルネットワークのあり得る制限とが明らかになることも認識した。例えば、特定の層の特定のニューロンは、特定の識別プロセスをさらに表すことができ、その結果、ニューラルネットワークプロセスのどの部分が不正確であるかに関してより正確に識別することができる。例えば、ニューラルネットワークが肺CT画像の潜在的な悪性結節の識別を実行するように構成されている場合、層のニューロンの第1のサブセットは結節の場所（悪性度の一因となる）の識別に関連し、一方、ニューロンの第2の異なるサブセットは結節の形状（同様に、悪性度の一因となる）の識別に関連する。したがって、ニューロンのサブセットのみ（例えば、潜在的な結節の場所に関連するもののみ）を使用することによって、より適切で有用な情報を得ることができる。

20

30

【0049】

特に、発明者等は、層のサブセットのみでなくその層のすべてのニューロンが使用される場合、かなりの量の情報が失われることを認識した。したがって、層のすべてのニューロンの活性化値を折り畳むと（例えば、比較のために単一の値に）、情報が失われる。

【0050】

ニューロンのサブセットのみを選択することにより、最も重要なニューロンのみ（例えば、ニューラルネットワーク入力の処理に最も貢献するもの）を識別することもできる。これにより、2つのニューラルネットワーク入力の活性化値のセットが類似しているか又は類似していないかを決定するのに、重要でないニューロンが過度に影響を与えるのを避けることができる。

40

【0051】

実施形態はまた、第1及び/又は第2のニューラルネットワーク入力なぜそのようなやり方で処理されたかをユーザが理解するのを助けるためにユーザに情報を提供するのに使用される。ユーザは、モデルがどのように第1のニューラルネットワーク入力と第2のニューラルネットワーク入力とを区別したかを調べるためにそれらの間の違いを探ることができる。これは、モデルの分類方法に関してユーザに一層の明確さをもたらし、それにより、モデルの確度及び信頼性へのユーザの信用が改善されるのを助ける。

【0052】

図2は、提案する概念が利用されているニューラルネットワーク1を示す。明確にする

50

ために、異なる層のニューロン間の接続は省略されている。

【 0 0 5 3 】

単一層 3 の 1 つ又は複数のニューロンのサブセット 3 1 が選択されている。サブセット 3 1 は、単一層に位置している一部のニューロン 3 a、3 bのみからなる（すなわち、ニューロンの層全体を含まない）。ニューロンがニューラルネットワーク入力进行处理しているときにニューロンのサブセットの活性化値を決定することによって、活性化値のセットが形成される。したがって、ニューラルネットワーク入力  $d_i$  は、ニューロン 3 a、3 b のサブセット 3 1 の活性化値のセット（図示せず）に関連づけられるとともに、これらのニューロン 3 a、3 b は、ニューラルネットワーク入力の処理の間にニューラルネットワーク入力  $d_i$  を処理している。

10

【 0 0 5 4 】

類似しているニューラルネットワーク出力に関連しているが、互いに類似していない活性化値のセットを有する（又はその逆である）2つのニューラルネットワーク入力を識別する目的が提案される。そのような識別により、ニューラルネットワーク入力の潜在的に誤った処理の識別が可能になり、それにより、ニューラルネットワーク出力が不正確であること、及び/又はニューラルネットワークで実行される処理が不正確である（例えば、未学習又は過学習である）ことが示される。

【 0 0 5 5 】

ニューロンのサブセット 3 1 を選択するための実施形態は後で説明される。しかしながら、ある実施形態は、ニューロンのどのサブセットが調査されるべきかを示すユーザ入力を受け取ることを含む。したがって、ニューラルネットワークの確度へのユーザ主導の調査があり得る。他の実施形態は、ニューロンの適切なサブセット 3 1 を自動的に選択する。

20

【 0 0 5 6 】

図 3 は、一実施形態によるデータを処理するためのコンピュータ実施方法 3 0 0 を示す。図示の方法 3 0 0 は、一般に、システムのプロセッサ（後で説明するものなど）の制御によって又はその制御下で実行される。この方法は、いくつかの実施形態に従って部分的に又は完全に自動化される。

【 0 0 5 7 】

方法 3 0 0 は、ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理し、それによって、第 1 のニューラルネットワーク出力を生成するステップ 3 0 1 を有する。方法 3 0 0 は、ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理し、それによって、第 2 のニューラルネットワーク出力を生成するステップ 3 0 2 をさらに有する。

30

【 0 0 5 8 】

このようにして、ニューラルネットワークの 2 つのニューラルネットワーク入力がニューラルネットワークによって処理されて、それぞれのニューラルネットワーク出力が提供される。ステップ 3 0 1 及び 3 0 2 は、並行して又は連続して実行される。ステップ 3 0 1 及び 3 0 2 が早い時点に実行され、ニューラルネットワーク出力は後の処理のために格納されてもよい（例えば、データベースに）ことが理解されよう。

【 0 0 5 9 】

方法 3 0 0 は、第 1 の類似性指標を決定するステップをさらに有し、第 1 の類似性指標は、第 1 のニューラルネットワーク出力と第 2 のニューラルネットワーク出力との間の類似性のレベルを示す。

40

【 0 0 6 0 】

第 1 の類似性指標は、例えば、第 1 のニューラルネットワーク出力と第 2 のニューラルネットワーク出力との間の（多次元）距離を示す。2つのニューラルネットワーク出力間の類似性を決定する他の方法は、当業者には明らかであろう。例えば、ニューラルネットワークが分類演算を実行する場合、第 1 の類似性指標は、第 1 及び第 2 のニューラルネットワーク入力と同じ分類に属するかを第 1 及び第 2 のニューラルネットワーク出力が示しているかどうかを示す。

50

## 【 0 0 6 1 】

方法 3 0 0 は、ニューラルネットワークの 1 つ又は複数のニューロンのセットを選択するステップ 3 0 4 をさらに有し、ニューロンのセットは、ニューラルネットワークの単一層のニューロンのサブセットである。したがって、同じ層 3 のニューロンのサブセット 3 1 が選択される。サブセットの選択は、例えば、ユーザ入力に基づくか、又は第 1 / 第 2 の入力ニューラルネットワーク入力の処理中のその層のニューロンの活性化値に基づく。層の選択は、ユーザ入力に基づいてもよく、又は所定の層（例えばニューラルネットワークの最外層）であってもよい。

## 【 0 0 6 2 】

特定の実施形態では、ステップ 3 0 4 は、ニューラルネットワークの単一のニューロンのみを選択するステップを有する。しかしながら、好ましい実施形態では、ステップ 3 0 4 は、同じ層の少なくとも 2 つのニューロンを選択するステップを有する。選択されたニューロンのサブセット 3 1 が単一層のすべてのニューロンを含むわけではないことは明らかであろう。

10

## 【 0 0 6 3 】

方法 3 0 0 は、ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理し、それによって、第 1 の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定するステップ 3 0 5 をさらに有する。方法 3 0 0 は、ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理し、それによって、第 2 の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定するステップ 3 0 6 をさらに有する。

20

## 【 0 0 6 4 】

このようにして、ステップ 3 0 5、3 0 6 は、ニューラルネットワークがそのニューラルネットワーク入力を処理している間にニューロンの選択されたサブセット 3 1 の活性化値を決定する（ニューラルネットワーク入力ごとに）。このようにして、活性化値の 2 つのセットが生成される。

## 【 0 0 6 5 】

この方法は、第 1 の活性化値のセットと対応する第 2 の活性化値のセットとの間の類似性のレベルを示す第 2 の類似性指標を決定するステップ 3 0 7 を有する。このようにして、2 つの活性化値のセットが比較されて、第 2 の類似性指標が生成され、それにより、2 つの間の類似性のレベルが示される。

30

## 【 0 0 6 6 】

ステップ 3 0 7 は、例えば、活性化値の 2 つのセットの間の多次元距離を決定するステップを有する。別の例では、ステップ 3 0 7 は、活性化値の 2 つのセットが所定の閾値（例えば、活性化値が 0 から 1 最大範囲を有する場合 0 . 5、すなわち 5 0 %）を超える同じ数の活性化値を共有するかどうかを決定するステップを有する。ステップ 3 0 7 は、セット間のいくつの対応する活性化値（例えば、第 1 のセットの第 1 の活性化値を第 2 のセットの第 1 の活性化値と比較する、第 1 のセットの第 2 の活性化値を第 2 のセットの第 2 の活性化値と比較する、など）が、互いに所定の許容誤差（例えば、 $\pm 1\%$  又は  $\pm 10\%$ ）内にあるかを決定するステップを有する。活性化値の 2 つのセット間の類似性を決定する様々な他の方法が、当業者には容易に明らかであろう。

40

## 【 0 0 6 7 】

活性化値の 2 つのセットは厳密に同一ではない（数値出力が連続スケールで生成されるので）ことが当業者には理解されよう。それゆえに、いくつかの実施形態では、第 2 の類似性指標は、活性化値のセットが互いに許容範囲（又は閾値類似性）内にあるかどうかを示す。

## 【 0 0 6 8 】

前に述べたように、いくつかの実施形態では、ステップ 3 0 7 は、第 1 の活性化値のセットと第 2 の活性化値のセットとの間の多次元距離を計算するステップを有する。例えば、各ニューロンを多次元空間における次元として考えると、ユークリッド距離を計算する

50

ことができる（例えば、第 1 の活性化値のセットと第 2 の活性化値のセットとの間の差の二乗和に従って）。このようにして、多次元距離を計算することができ、その距離は、活性化値の 2 つのセット間の類似性を示す。

【 0 0 6 9 】

当然、ニューロンのサブセット 3 1 が単一のニューロンしか含まない場合、ステップ 3 0 7 は、第 1 の活性化値のセットの単一の活性化値と第 2 の活性化値のセットの単一の活性化値との間の数値差を決定するステップを有する。

【 0 0 7 0 】

いくつかの実施形態では、ステップ 3 0 7 は、第 1 の活性化値のセットの統計的に有意な活性化値の数と第 2 の活性化値のセットの統計的に有意な活性化値の数との間の差を識別するステップを有する。ここで、統計的に有意な活性化値は、そのニューロンの平均活性化値からの特定の数（例えば、2、3、又は 4）の標準偏差を超えているものである。他の実施形態では、統計的に有意な活性化値は、活性化値の所定の範囲の外にあるものである。

10

【 0 0 7 1 】

次いで、方法 3 0 0 は、第 1 の類似性指標が第 1 の所定の閾値を超える類似性のレベルを示しており、第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示している（又は逆である）かどうかを決定するステップ 3 0 8 を有する。

【 0 0 7 2 】

したがって、ステップ 3 0 8 は、代替として又はさらに、第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示しており、第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示しているかどうかを決定するステップを有する。

20

【 0 0 7 3 】

所定の閾値の値及び単位は、当然、類似性指標の性質に依存する。例えば、類似性指標が多次元距離によって表される場合、所定の閾値は所定の距離である。

【 0 0 7 4 】

このようにして、ステップ 3 0 8 は、効果的に排他的 O R 関数（X O R）を実行して、第 1 及び第 2 のニューラルネットワーク出力が類似しているが、2 つの対応する活性化値のセットが類似していない（又は逆である）かどうかを決定する。

【 0 0 7 5 】

30

このようにして、第 1 及び第 2 のニューラルネットワーク入力の中の 1 つが、ニューラルネットワークによって不正確に処理されたかどうかに関して予測を行うことができる。ニューラルネットワークは類似しているニューラルネットワーク出力に関連するニューラルネットワーク入力を同様のやり方で処理すべきであり、その結果、この仮定からの逸脱はニューラルネットワークによる潜在的な不正確な処理を示すことが認識されているので、この予測を行うことができる。

【 0 0 7 6 】

そのような認識により、ニューラルネットワーク入力の潜在的な不正確処理を評価し考慮に入れることができる。例えば、潜在的な不正確がユーザにフラグで警告され（例えば、手動評価の場合）、手動評価の結果はニューラルネットワークを再訓練するために（すなわち、フィードバックシステムとして）使用される。潜在的な不正確はまた、ユーザがニューラルネットワークの基礎をなすプロセスを理解するのを助けるためにユーザにフラグで警告されることもある。いくつかの例では、潜在的な不正確の識別により、例えば、ユーザフィードバックに基づく又は新しい訓練データの取得によるニューラルネットワークの再訓練が促される。

40

【 0 0 7 7 】

実施形態は、分類プロセスを実行するように構成されたニューラルネットワークで利用される場合に（すなわち、ニューラルネットワーク出力がニューラルネットワーク入力の分類である場合に）特に有利である。この理由は、不正確な分類（例えば、医用画像の）がニューラルネットワーク入力に関しての誤診又は誤解をもたらすことがあるからである

50

。このように、潜在的な誤分類を識別する特別な動機がある。

【 0 0 7 8 】

したがって、いくつかの実施形態では、ステップ 3 0 1 及び 3 0 2 は、ニューラルネットワークを使用してそれぞれ第 1 及び第 2 のニューラルネットワーク入力の分類を決定するステップを有する。対応するステップ 3 0 3 は、第 1 及び第 2 のニューラルネットワーク入力と同じ分類に分類されるかどうか（すなわち、対応する第 1 及び第 2 のニューラルネットワーク出力が同じ分類を示すかどうか）を示す第 1 の類似性指標を生成するステップを有する。ステップ 3 0 8 では、第 1 の類似性指標が、第 1 及び第 2 のニューラルネットワーク出力（それによって、ニューラルネットワーク入力）が同じ分類のものであることを示す場合、第 1 の類似性指標は第 1 の所定の閾値を超える類似性のレベルを示すことが仮定されている。同様に、第 1 の類似性指標が、第 1 及び第 2 のニューラルネットワーク出力（それによって、ニューラルネットワーク入力）が異なる分類のものであることを示す場合、第 1 の類似性指標は第 3 の所定の閾値未満の類似性のレベルを示すことが仮定されている。

10

【 0 0 7 9 】

いくつかの実施形態では、図 3 のステップ 3 0 4 は、ニューロンの前記第 1 のセットに関連する所定の活性化範囲の外の第 1 のニューラルネットワーク入力によって活性化される（ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理するブロック 3 0 1 の間に）ニューラルネットワークの単一層のニューロンのサブセットを識別するステップを有する。したがって、ニューロンのサブセットの選択は、第 1 のニューラルネットワーク入力の処理の間に前記ニューロンの活性化値が所定の活性化範囲の外にあるニューロン（同じ層の）を識別することを含む。

20

【 0 0 8 0 】

より詳細には、上述のように、層の各ニューロンは、上下限範囲内で活性化される（例えば、数値出力を生成する）。活性化値の上下限範囲は、ニューロンによって実行される数学的演算に依存する。これらの上下限範囲内で、ニューロンは、上下限内の典型的な範囲（例えば統計的に正常な範囲）を有する。例えば、ニューロンは上下限範囲間で活性化されるが、大部分のニューラルネットワーク入力に対して、ニューロンは、ニューロンに関連する活性化範囲内でのみ活性化される。ニューラルネットワーク入力を分類するとき、分類は、ネットワークの特定のニューロンにおいて強力な（例えば、統計的に異常な）活性化を生成し、それによって、活性化がニューロンに関連する活性化範囲の外にある（超えている / 未満である）ニューラルネットワーク入力の特定の強力な特性又は特徴に基づいて、すなわち、統計的に外にある活性化値を識別して、決定される。

30

【 0 0 8 1 】

そのため、実施形態では、ニューロンの前記第 1 のセットに関連する活性化範囲の外に活性化されたニューラルネットワークのニューロンのセットを識別することには、ニューロンの活性化値（すなわち、第 1 のニューラルネットワーク入力を分類するのに使用される活性化値）を第 2 のニューラルネットワーク入力を分類するのに使用される同じニューロンの活性化値と比較することが含まれる。例えば、活性化値の分布の平均活性化値及び関連する標準偏差が計算される。したがって、ニューロンの前記セットに関連する活性化範囲の外に活性化されたニューラルネットワークのニューロンのセットを識別することには、ニューロンの平均活性化から（例えば、上に又は下に）特定の数の標準偏差を超えて活性化される（第 1 のニューラルネットワーク入力によって）単一層のニューロンを識別することが含まれる。

40

【 0 0 8 2 】

いくつかの実施形態では、ニューロンの前記第 1 のセットに関連する活性化範囲の外に活性化されたニューラルネットワークのニューロンの第 1 のセットを識別することには、ニューロンの活性化値（第 1 のニューラルネットワーク入力を分類するときの）を第 1 のニューラルネットワーク入力を分類するときの同じ層の他のニューロンの活性化値と比較することが含まれる。一般に、同じ層のニューロンは同じ数学的演算（異なる重み付けを

50



もつ)を含み、したがって、層のニューロンの活性化は、直接比較することができる(例えば、シグモイド層は、常に、0と1との間の出力を生成する)。したがって、層内のニューロンの活性化値は、層において最も強力に(例えば、統計的に異常に)活性化されたニューロンを見いだすために比較される。

【0083】

提案したステップ304は、代わりに、第2のニューラルネットワーク入力に基づいて単一層のニューロンのサブセット31を識別する(例えば、第2のニューラルネットワーク入力を処理するとき所定の範囲の外の活性化値を有するニューロンを識別する)ように、必要な変更を加えて、変形される。

【0084】

いくつかの実施形態では、ステップ304は、ニューラルネットワークの外層のニューロンのセットを選択するステップを有する。ここで、外層は、ニューラル層の出力(例えば、出力ノード $n_o$ )に最も近い層のうちの1つを指す。外層は、例えば、ニューラルネットワークの最終の3分の1の層のうちの1つである。好ましくは、外層は、最外層4(例えば、出力 $n_o$ の前のニューラルネットワークの最終層4)である。外層は、ニューラルネットワーク入力の最大の特徴(画像中の対象全体など)によって活性化され、そのため、ニューラルネットワークの外層に共通の活性化値を有するニューラルネットワーク入力のニューロンのサブセットは、最も類似している全体的特徴を有するニューラルネットワーク入力に対応する。そのため、ニューラルネットワークの外層のニューロンのサブセットにおいて類似している活性化値を有するが、しかしなお類似していないニューラルネットワーク出力を有する(又は逆である)ニューラルネットワーク入力を識別することには、類似してはいるが、それにもかかわらず、違ったやり方で処理された(又は逆である)大規模な特徴をもつニューラルネットワーク入力を識別するという利点がある。これにより、潜在的に誤って処理されたニューラルネットワーク入力を識別する際の確度が改善される。

【0085】

上位層のニューロンのサブセットを使用することに関連した利点があるが、任意の層のニューロンのサブセットを使用できることを当業者なら理解されよう。

【0086】

いくつかの実施形態では、ステップ307が第1の活性化値のセットと第2の活性化値のセットとの間の多次元距離を決定するステップを有する場合、ステップ308の第2及び/又は第4の所定の閾値は、所定の閾値距離を含む。当業者なら理解するように、ステップ308で使用するための適切な閾値は、ニューラルネットワークのニューロンと、処理されるニューラルネットワーク入力とに依存する。

【0087】

方法300は、複数の異なる第2のニューラルネットワーク入力に対して実行される。すなわち、方法のステップは、第1のニューラルネットワーク入力、1つを超える第2のニューラルネットワーク入力と比較されるように変更される。しかしながら、好ましい実施形態では、ニューロンの選択されたセットは、第1のニューラルネットワーク入力及び各第2のニューラルネットワーク入力に対して同じである。言い換えれば、方法300は、第1のニューラルネットワーク入力を1つを超える第2のニューラルネットワーク入力と比較するように構成される。これは、ステップ308の要件を満たす第2のニューラルネットワーク入力識別されるのを可能にし、それによって、誤って処理された可能性があるさらなるニューラルネットワーク入力の識別を可能にする。

【0088】

言い換えれば、ステップ302は、ニューラルネットワークを使用して複数の第2のニューラルネットワーク入力を処理し、それによって、複数の第2のニューラルネットワーク出力を生成するステップを有するように構成される。ステップ303は、複数の第2のニューラルネットワーク入力の各々に対して、第1のニューラルネットワーク出力と第2のニューラルネットワーク出力との間の類似性のレベルを示す第1の類似性指標を決定し

10

20

30

40

50

、それによって、複数の第 1 の類似性指標を生成するステップを有するように構成される。ステップ 306 は、複数の第 2 のニューラルネットワーク入力 of 各々に対して、ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理するときに各選択されたニューロンの活性化値を決定し、それによって、第 2 のニューラルネットワーク入力ごとに第 2 の活性化値のセットを生成し、それによって、複数の第 2 の活性化値のセットを生成するステップを有するように構成される。ステップ 307 は、複数の第 2 のニューラルネットワーク入力の各々に対して、第 1 の活性化値のセットと対応する第 2 の活性化値のセットとの間の類似性のレベルを示す第 2 の類似性指標を決定し、それによって、複数の第 2 の類似性指標を生成するステップを有するように構成される。ステップ 308 は、第 2 のニューラルネットワーク入力ごとに、第 1 の類似性指標が第 1 の所定の閾値を超える類似性のレベルを示し、第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示すか、又は第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示し、第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示すかどうかを決定し、それによって、第 2 のニューラルネットワーク入力ごとに、第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つがニューラルネットワークによって不正確に処理されたかどうかを予測するステップを有するように構成される。ステップ 308 は、第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つがニューラルネットワークによって不正確に処理されたことが予測されるすべての第 2 のニューラルネットワーク入力を識別するステップをさらに有する。

10

【0089】

20

このようにして、想定される方法は、類似しているニューラルネットワーク出力に関連するが、類似していない活性化値のセットに関連する、又はその逆に関連する 1 つ又は複数の第 2 のニューラルネットワーク入力を識別する。これにより、潜在的に誤って処理されたニューラルネットワーク入力のデータセットを識別することができ、それがユーザに提供されて、ニューラルネットワークの基礎をなすプロセスをユーザが理解すること、又はニューラルネットワークの再訓練をガイドすることを助ける。

【0090】

いくつかの実施形態では、ステップ 307 は、第 1 の活性化値のセット及び各第 2 の活性化値のセットに多次元クラスタリングを実行するステップを有し、その結果、第 2 の類似性指標は、第 1 の活性化値のセットが、対応する第 2 の活性化値のセットと同じクラスタにあるかどうかを示す。例えば、第 1 のニューラルネットワーク入力に関連する第 1 の活性化値のセット、及び複数の第 2 のニューラルネットワーク入力の各々に関連する第 2 の活性化値のセットが、多次元パラメータ空間にプロットされる。そのような実施形態では、類似性指標は、第 1 及び第 2 の活性化値のセットが多次元パラメータ空間において同じクラスタに配置されているかどうかを示す。

30

【0091】

対応するステップ 308 は、第 1 のニューラルネットワーク入力の第 1 の活性化値のセットに関連する少なくとも 1 つのクラスタ内にあり、第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示す第 2 の活性化値のセットに関連する第 2 のニューラルネットワーク入力を識別するステップを有する。代替として及び / 又は追加として、ステップ 308 は、第 1 のニューラルネットワーク入力の第 1 の活性化値のセットに関連する少なくとも 1 つのクラスタ内にあり、第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示す第 2 の活性化値のセットに関連する第 2 のニューラルネットワーク入力を識別するステップを有する。

40

【0092】

方法 300 が第 1 のニューラルネットワーク入力の異なるインスタンスに対して反復して繰り返されてもよいことを当業者は理解されよう。例えば、ニューラルネットワーク入力のデータベースがあり、ニューラルネットワーク入力の各々が、方法 300 に従って処理される（例えば、ニューラルネットワーク入力の各々が、データベースの他の各々のニューラルネットワーク入力と比較される）。これにより、潜在的に誤って処理されたニュー

50

ーラルネットワーク入力を識別する可能性が高まる。

【 0 0 9 3 】

当業者はまた、ニューロンの異なるサブセットに対して方法 3 0 0 を繰り返すことができることを認識するであろう。そのような事例では、ステップ 3 0 1 ~ 3 0 3 を繰り返す必要がない（ニューラルネットワーク出力間の類似性が既に識別されているので）。したがって、ステップ 3 0 4 ~ 3 0 8 が、ニューロンの異なるサブセットに対して繰り返される。これにより、さらに多くの追加の情報が確認される。

【 0 0 9 4 】

上述のいくつかの実施形態によれば、第 1 のニューラルネットワーク入力と類似性（活性化値のセット間又はニューラルネットワーク出力間のいずれか一方の）を共有するが、非類似性（活性化値のセット及びニューラルネットワーク出力の他方の）も共有する少なくとも 1 つの第 2 のニューラルネットワーク入力を識別するための方法及びシステムがある。例えば、識別された第 2 のニューラルネットワーク入力、第 1 のニューラルネットワーク入力と共通の活性化値のセットを有するニューラルネットワーク入力を表すことができるので、第 1 のニューラルネットワーク入力とある意味で類似していると思なされるが、それにもかかわらず、ニューラルネットワークは、ニューラルネットワーク入力に関連するものとは類似していないニューラルネットワーク出力を提供する。それゆえに、方法 3 0 0 は、ニューラルネットワークの基礎をなすプロセスの解明を助けるのに使用することができる特定のニューラルネットワーク入力を識別するために使用される。

【 0 0 9 5 】

いくつかの実施形態では、識別された第 2 のニューラルネットワーク入力は反例であると思なされる。その理由は、それが、第 1 のニューラルネットワーク入力に対して類似しているニューラルネットワーク入力で生じた反例を示すからである。このようにして、この特定の第 2 のニューラルネットワーク入力は、モデルが、第 1 のニューラルネットワーク入力を処理したやり方、及び / 又は第 1 のニューラルネットワーク入力に与えられたニューラルネットワーク出力（例えば、分類）に類似しているニューラルネットワーク出力（例えば、分類）間を区別するやり方をユーザに強調表示するために使用される。したがって、不正確に処理されたニューラルネットワーク入力を認識するのではなく、この方法は、ニューラルネットワークシステムの処理を解明するために使用される。

【 0 0 9 6 】

図 3 に示されていないが、本明細書に記載された実施形態のいずれも、識別された第 2 のニューラルネットワーク入力を訓練プロセスで使用してニューラルネットワークを改善することをさらに含むことができる。

【 0 0 9 7 】

例えば、識別された第 2 のニューラルネットワーク入力をユーザに表示することを含む実施形態では、ユーザは、各第 2 のニューラルネットワーク入力及び / 又は第 1 のニューラルネットワーク入力に関するフィードバックを行うことができる。例えば、ユーザは、特定のニューラルネットワーク入力及び / 又は第 1 のニューラルネットワーク入力に対する正しいニューラルネットワーク出力に関するフィードバックを行うことができる。そのようなフィードバックは、モデルを再訓練するために使用される。例えば、フィードバックは、モデルをさらに訓練するためのニューラルネットワーク入力（例えば、識別された第 2 のニューラルネットワーク入力に類似している又は類似していない）を選択するために使用される。

【 0 0 9 8 】

代替として又は追加として、ユーザは、ユーザがモデルの基礎をなすプロセスをよりよく理解するのを第 2 のニューラルネットワーク入力が助けたかどうかに関してフィードバックを行うことができる。そのようなフィードバックは、ニューラルネットワークが基礎をなすプロセスに対する信用をさらに改善するためにユーザに提供される情報を改善するように第 2 のニューラルネットワーク入力の選択を改善するのに使用される。

【 0 0 9 9 】

図3に示されていないが、本明細書に記載された実施形態のいずれも、ニューラルネットワークを使用して、第1のニューラルネットワーク入力に類似している第2のニューラルネットワーク入力を（例えば、複数の第2のニューラルネットワーク入力から）識別すること、すなわち、第1の類似性指標（第1及び第2のニューラルネットワーク出力間の類似性を示す）が第1の所定の閾値を超える類似性を示し、第2の類似性指標（同じ層のニューロンの同じサブセットの活性化値の2つのセットの間の類似性を示す）が第3の所定の閾値を超える類似性を示す（すなわち、類似している）場合を識別することをさらに含む。いくつかの実施形態では、この方法は、類似している第2のニューラルネットワーク入力をユーザに表示するステップをさらに有する。このようにして、ユーザは、例えば、（例えば、類似している活性化値のセットを有する）ように異なるやり方で処理され、類似している出力を有する第2のニューラルネットワーク入力に加えて、（例えば、類似している活性化値のセットを有する）ように同様のやり方で処理され、第1のニューラルネットワーク入力に対して類似している出力（例えば、同一の分類）を有する（又は逆である）第2のニューラルネットワーク入力の1つ又は複数の例が提供される。これは、第1のニューラルネットワーク入力と同じやり方で処理された第2のニューラルネットワーク入力のタイプ、並びに第1のニューラルネットワーク入力と違うように処理された第2のニューラルネットワーク入力のタイプに対するさらなる洞察を提供する。

#### 【0100】

したがって、提案する実施形態はまた、第2のニューラルネットワーク入力、第1のニューラルネットワーク入力に関連する活性化値のセット及びニューラルネットワーク出力と同様である活性化値のセット及びニューラルネットワーク出力に関連しているかどうかを識別することができる。このようにして、第1/第2のニューラルネットワーク入力

#### 【0101】

が正確に処理されたかどうかに関して予測を行うことができる。図3に示されていないが、本明細書に記載された実施形態のいずれも、例えば、以下で説明するシステム500のユーザインタフェース504を使用して第2のニューラルネットワーク入力をユーザに表示することをさらに含む。いくつかの実施形態では、第1のニューラルネットワーク入力は、第2のニューラルネットワーク入力と一緒にユーザに表示される。

#### 【0102】

いくつかの実施形態では、ニューラルネットワークが分類演算を実行するとき、第1のニューラルネットワーク入力の分類及び/又は第2のニューラルネットワーク入力の分類がユーザにさらに表示される。このようにして、ユーザは、第1のニューラルネットワーク入力を第2のニューラルネットワーク入力と比較し、ニューラルネットワークの分類プロセスに対するより深い洞察を得ることができる。例えば、ユーザは、モデルに第1の/第2のニューラルネットワーク入力を分類させた第1のニューラルネットワーク入力と第2のニューラルネットワーク入力との間の類似性及び/又は差異を、モデルが行ったように観察することができる。

#### 【0103】

実施形態では、第2のニューラルネットワーク入力に関する他の情報、例えば、第2のニューラルネットワーク入力についての1つ又は複数の注釈などが、ユーザに提供される。注釈は、ユーザ（例えば、医療専門家など）によって加えられたニューラルネットワーク入力に関する情報である。表示された注釈がニューラルネットワークを訓練するために使用される場合、そのような注釈は、ニューラルネットワークの基礎をなすプロセスに対する洞察をさらに提供し、ニューラルネットワークを改善するために使用される（例えば、訓練プロセスにおいてニューラルネットワークに提供される注釈の品質を改善することによって）。

#### 【0104】

注釈の例には、限定はしないが、ニューラルネットワーク入力の特徴のユーザ分類又はラベル（例えば、ユーザが、解剖学的構造の場所又は異常の場所などのニューラルネット

10

20

30

40

50

ワーク入力の特徴を強調表示する)と、例えば、ニューラルネットワーク入力の内容から導き出されるような症状又は病気についてのユーザの診断などのニューラルネットワーク入力の内容のユーザ解釈とが含まれる。

【0105】

図4は、一実施形態による方法400の例示の出力を示し、結果はユーザに表示される。この実施形態では、第1のニューラルネットワーク入力は先端の尖った形状の画像402を含み、それは、例えば、医用画像の解剖学的対象物を表す。3つの第2のニューラルネットワーク入力404、406、408が、複数の第2のニューラルネットワーク入力処理され第1のニューラルネットワーク入力と比較される上述の図3に関して説明した方法に従って識別された。第2のニューラルネットワーク入力(ステップ308で述べた要件を満たす)が、第1のニューラルネットワーク入力の隣に表示される。

10

【0106】

図示の実施形態では、各識別された第2のニューラルネットワーク入力は、第1の活性化値のセットに対して類似していない活性化値のセット(単一層のニューロンの識別されたサブセットについて)に関連するが、第1のニューラルネットワーク出力に対して類似している第2のニューラルネットワーク出力に関連する。他のあり得る識別された第2のニューラルネットワーク入力は、各々、類似している第2の活性化値のセット(第1の活性化値のセットに対して)に関連するが、類似していない第2のニューラルネットワーク出力(第1のニューラルネットワーク出力に対して)に関連することが理解されよう。

【0107】

20

3つの「類似しているニューラルネットワーク入力」が、さらに、410、412、414で表示されている。図示の実施形態では、各識別された第2のニューラルネットワーク入力は、第1の活性化値のセット(第1のニューラルネットワーク入力402に関連する)と同様である第2の活性化値のセット(単一層のニューロンの識別されたサブセットでの)に関連するが、第1のニューラルネットワーク入力402の第1のニューラルネットワーク出力に対して類似していない第2のニューラルネットワーク出力に関連する。

【0108】

この例では、ニューラルネットワーク入力の一部についての特徴の一部の注釈416が、さらに、追加のコンテキストデータ(例えば、各ニューラルネットワーク入力に関連する対象者の年齢)と一緒に表示されている。

30

【0109】

図5は、ニューラルネットワークの入力データを処理するために使用される一実施形態によるシステム500のブロック図を示す。図5を参照すると、システム500はプロセッサ502を含み、プロセッサ502は、システム500の動作を制御し、本明細書に記載された方法を実施する。システム500は、命令のセットを表す命令データを含むメモリ506をさらに含む。メモリ506は、本明細書に記載された方法を実行するためにプロセッサ502によって実行されるプログラムコードの形態の命令データを格納するように構成される。いくつかの実施態様では、命令データは、各々が本明細書に記載された方法の個々のステップ又は多数のステップを実行するように構成されるか又は実行するためのものである複数のソフトウェア及び/又はハードウェアモジュールを含む。いくつかの実施形態では、メモリ506は、システム500の1つ又は複数の他の構成要素(例えば、システム500のプロセッサ502及び/又は1つ又は複数の他の構成要素)をさらに含むデバイスの一部である。代替実施形態では、メモリ506は、システム500の他の構成要素とは別個のデバイスの一部である。

40

【0110】

システム500のプロセッサ502は、命令のセットを実行するためにメモリ506と通信するように構成される。命令のセットは、プロセッサによって実行されると、本明細書に記載された方法をプロセッサに実行させる。プロセッサ502は、本明細書に記載されたやり方でシステム500を制御するように構成又はプログラムされた1つ又は複数のプロセッサ、処理ユニット、マルチコアプロセッサ、又はモジュールを含む。いくつかの

50

実施態様では、例えば、プロセッサ 5 0 2 は、分散処理のために構成された複数のプロセッサ、処理ユニット、マルチコアプロセッサ、及び / 又はモジュールを含む。そのようなプロセッサ、処理ユニット、マルチコアプロセッサ、及び / 又はモジュールは、異なる場所に配置されてもよく、各々が本明細書に記載された方法の異なるステップ及び / 又は単一ステップの異なる部分を実行してもよいことを当業者なら理解されよう。

#### 【 0 1 1 1 】

簡単に言えば、命令のセットは、システム 5 0 0 のプロセッサ 5 0 2 によって実行されると、ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理し、それによって、第 1 のニューラルネットワーク出力を生成すること、ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理し、それによって、第 2 のニューラルネットワーク出力を生成すること、第 1 の類似性指標を決定することであり、第 1 の類似性指標が第 1 のニューラルネットワーク出力と第 2 のニューラルネットワーク出力との間の類似性のレベルを示す、決定すること、ニューラルネットワークの 1 つ又は複数のニューロンのセットを選択することであり、ニューロンのセットがニューラルネットワークの単一層のニューロンのサブセットである、選択すること、ニューラルネットワークを使用して第 1 のニューラルネットワーク入力を処理し、それによって、第 1 の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定すること、ニューラルネットワークを使用して第 2 のニューラルネットワーク入力を処理し、それによって、第 2 の活性化値のセットを生成するときに、各選択されたニューロンの活性化値を決定すること、第 1 の活性化値のセットと対応する第 2 の活性化値のセットとの間の類似性のレベルを示す第 2 の類似性指標を決定すること、第 1 の類似性指標が第 1 の所定の閾値を超える類似性のレベルを示し、第 2 の類似性指標が第 2 の所定の閾値未満の類似性のレベルを示すか、又は第 1 の類似性指標が第 3 の所定の閾値未満の類似性のレベルを示し、第 2 の類似性指標が第 4 の所定の閾値を超える類似性のレベルを示すかどうかを決定し、それによって、第 1 及び第 2 のニューラルネットワーク入力のうちの 1 つがニューラルネットワークによって不正確に処理されたかどうかを予測することをプロセッサ 5 0 2 に行わせる。

#### 【 0 1 1 2 】

いくつかの実施形態では、命令のセットはまた、プロセッサ 5 0 2 によって実行されると、本明細書に記載された方法に関連するデータ及び情報を格納するためにプロセッサ 5 0 2 にメモリ 5 0 6 を制御させる。例えば、メモリ 5 0 6 は、本明細書に記載された方法に由来する、ニューラルネットワーク、第 1 のニューラルネットワーク入力、第 2 のニューラルネットワーク入力、ニューラルネットワーク出力、識別された第 2 のニューラルネットワーク入力、又は任意の他のデータ若しくは情報、又はデータ及び情報の任意の組合せのうちのいずれかを格納するために使用される。

#### 【 0 1 1 3 】

本明細書に記載された実施形態のいずれにおいても、第 1 及び / 又は第 2 のニューラルネットワーク入力は、ニューラルネットワーク出力を生成するためにニューラルネットワークで処理できるデータを含む。例えば、第 1 及び / 又は第 2 のニューラルネットワーク入力は、テキスト、画像データ、センサデータ、機器ログ、及び / 又は記録のうちの任意の 1 つ又はそれらの組合せを含む。いくつかの実施形態では、第 1 及び / 又は第 2 のニューラルネットワーク入力は、医用画像（例えば、CT スキャン、X 線スキャン、又は任意の他の適切な医用イメージング方法から取得された画像）、医療機器又はセンサ（心拍数モニタ、血圧モニタ、又は他のモニタなど）からの出力、又は医療記録のうちの任意の 1 つ又はそれらの任意の組合せなどの医療データを含む。したがって、ニューラルネットワーク入力は、ニューラルネットワークで処理できる任意のデータ（例えば、限定はしないが、医用画像、医療記録、及び / 又はセンサ読み取り値、又はニューラルネットワークで処理される任意の他のタイプのデータを含む画像、値、及び / 又は文字数字の列の任意の組合せ）を含む。様々なタイプのニューラルネットワーク入力の例が提供されたが、当業者は、本明細書で提供される教示が、ニューラルネットワークで処理できる任意の他のタ

イプのデータに等しく適用されることを理解されよう。

【 0 1 1 4 】

いくつかの実施形態では、図 5 に示すように、システム 5 0 0 は、本明細書に記載されたユーザ入力のいずれかを受け取るように構成された少なくとも 1 つのユーザインタフェース 5 0 4 を含む。ユーザインタフェース 5 0 4 は、システム 5 0 0 のユーザが、本明細書に記載された方法に関連する命令、データ、フィードバック、又は情報を手動で入力できるようにする。例えば、ユーザインタフェース 5 0 4 は、本明細書に記載された方法を実行する（例えば、分類プロセスを実行する）第 1 及び／又は第 2 のニューラルネットワーク入力を選択又は示すためにユーザによって使用される。ユーザインタフェース 5 0 4 は、システム 5 0 0 のユーザがユーザ入力を供給する、システム 5 0 0 と対話する、及び／又はシステム 5 0 0 を制御するのを可能にする任意のタイプのユーザインタフェースである。例えば、ユーザインタフェース 5 0 4 は、ユーザが、分類すべき第 1 のニューラルネットワーク入力を選択することができる 1 つ又は複数のスイッチ、1 つ又は複数のボタン、キーパッド、キーボード、マウス、タッチスクリーン若しくはアプリケーション（例えば、タブレット若しくはスマートフォンによる）、任意の他のユーザインタフェース、又はユーザインタフェースの組合せを含む。

10

【 0 1 1 5 】

いくつかの実施形態では、ユーザインタフェース 5 0 4（又はシステム 5 0 0 の別のユーザインタフェース）は、システム 5 0 0 のユーザへの情報、データ、又は信号のレンダリング（又は出力若しくは表示）を可能にする。そのため、ユーザインタフェース 5 0 4 は、システム 5 0 0 のユーザ（例えば、医療専門家、ヘルスケア提供者、ヘルスケア専門家、介護者、対象者、又は任意の他のユーザ）に本明細書の実施形態による方法に関連する又は由来する情報を提供する際に使用するためのものである。プロセッサ 5 0 2 は、本明細書の実施形態による方法に由来する情報を提供するために 1 つ又は複数のユーザインタフェース 5 0 4 を制御するように構成される。例えば、プロセッサ 5 0 2 は、本明細書に記載された方法に由来する第 1 及び／又は第 2 のニューラルネットワーク入力又は任意の他の情報（出力など）、或いは任意の組合せ情報をレンダリングする（又は出力若しくは表示する）ために 1 つ又は複数のユーザインタフェース 5 0 4 を制御するように構成される。ユーザインタフェース 5 0 4 は、いくつかの実施形態では、表示スクリーン、グラフィカルユーザインタフェース（GUI）若しくは他のビジュアルレンダリングコンポーネント、1 つ又は複数のスピーカ、1 つ又は複数のマイクロホン若しくは任意の他のオーディオコンポーネント、1 つ又は複数のライト、触覚フィードバック（例えば、振動機能）を提供するための構成要素、任意の他のユーザインタフェース、又は方法に関連若しくは由来する情報をユーザに提供するためのユーザインタフェースの組合せを含む。いくつかの実施形態では、ユーザインタフェース 5 0 4 は、システム 5 0 0 の 1 つ又は複数の他の構成要素（例えば、システム 5 0 0 のプロセッサ 5 0 2、メモリ 5 0 6、及び／又は 1 つ又は複数の他の構成要素）をさらに含むデバイスの一部である。代替実施形態では、ユーザインタフェース 5 0 4 は、システム 5 0 0 の他の構成要素とは別個のデバイスの一部である。

20

30

【 0 1 1 6 】

いくつかの実施形態では、図 5 に示すように、システム 5 0 0 は、システム 5 0 0 がシステム 5 0 0 の内部又は外部にある任意のインタフェース、メモリ、及びデバイスと通信できるようにするための通信用インタフェース（又は回路）5 0 8 をさらに含む。通信用インタフェース 5 0 8 は、任意のインタフェース、メモリ、及びデバイスと無線で又は有線接続を介して通信する。

40

【 0 1 1 7 】

図 5 は、本開示のこの態様を示すのに必要な構成要素を示しており、実際の実施態様では、システム 5 0 0 は図示のものに加えて他の構成要素を含むことが理解されよう。例えば、システム 5 0 0 は、システム 5 0 0 に電力を供給するためのバッテリー若しくは他の電源、又はシステム 5 0 0 を主電源に接続するための手段を含む。

50

## 【 0 1 1 8 】

コンピュータ可読媒体を含むコンピュータプログラム製品がさらに提供され、コンピュータ可読媒体は、その中に具現されたコンピュータ可読コードを有し、コンピュータ可読コードは、適切なコンピュータ又はプロセッサによる実行により、コンピュータ又はプロセッサに本明細書に記載された1つ又は複数の方法を実行させるように構成される。したがって、本開示はまた、実施形態を実行するように構成されたプログラム、特に、キャリア上又はキャリア内のコンピュータプログラムに適用されることが理解されよう。プログラムは、ソースコード、オブジェクトコード、部分的にコンパイルされた形態などのコード中間ソース及びオブジェクトコード、又は本明細書に記載された実施形態による方法の実施で使用するのに適する任意の他の形態とすることができる。

10

## 【 0 1 1 9 】

そのようなプログラムは多くの異なるアーキテクチャ設計を有してもよいことも理解されよう。例えば、方法又はシステムの機能を実施するプログラムコードは、1つ又は複数のサブルーチンに細分されてもよい。これらのサブルーチン間に機能を分散させる多くの異なるやり方が当業者には明らかであろう。サブルーチンを1つの実行可能ファイルと一緒に格納して、自己完結型プログラムを形成することができる。そのような実行可能ファイルは、コンピュータ実行可能命令、例えば、プロセッサ命令及び/又はインタープリタ命令(例えば、Javaインタープリタ命令)を含む。代替として、サブルーチンの1つ又は複数或いはすべては、少なくとも1つの外部ライブラリファイルに格納され、例えば、実行時に、主プログラムに静的又は動的にリンクされてもよい。主プログラムは、サブルーチンのうちの少なくとも1つへの少なくとも1つの呼出しを含む。サブルーチンはまた、互いの機能呼出しを含んでもよい。

20

## 【 0 1 2 0 】

コンピュータプログラム製品に関連する一実施形態は、本明細書に記載されている方法のうちの少なくとも1つの各処理段階に対応するコンピュータ実行可能命令を含む。これらの命令は、サブルーチンに細分され、及び/又は静的又は動的にリンクされる1つ又は複数のファイルに格納されてもよい。コンピュータプログラム製品に関連する別の実施形態は、本明細書に記載されているシステム及び/又は製品のうちの少なくとも1つの各手段に対応するコンピュータ実行可能命令を含む。これらの命令は、サブルーチンに細分され、及び/又は静的又は動的にリンクされる1つ又は複数のファイルに格納されてもよい。

30

## 【 0 1 2 1 】

コンピュータプログラムのキャリアは、プログラムを搬送できるエンティティ又はデバイスとすることができる。例えば、キャリアは、ROM、例えばCD ROM若しくは半導体ROM、又は磁気記録媒体、例えばハードディスクなどのデータストレージを含むことができる。さらに、キャリアは、電気ケーブル若しくは光ケーブルを介して、又は無線若しくは他の手段によって移送される電気又は光信号などの伝達可能なキャリアとすることができる。プログラムがそのような信号で具現される場合、キャリアは、そのようなケーブル又は他のデバイス若しくは手段で構成することができる。代替として、キャリアは、プログラムが埋め込まれる集積回路とすることができ、集積回路は、関連する方法を実行するように構成されるか、又は関連する方法の実行の際に使用される。

40

## 【 0 1 2 2 】

図面、開示、及び添付の特許請求の範囲の検討から、開示された実施形態に対する変形が、当業者によって理解され達成される。特許請求の範囲において、「有する、備える、含む」という単語は、他の要素又はステップを排除せず、単数形は、複数を除外しない。単一のプロセッサ又は他のユニットは、特許請求の範囲に列挙されているいくつかの項目の機能を遂行することができる。特定の手段が相互に異なる従属請求項に列挙されているという単なる事実は、これらの手段の組合せを有利に使用できないことを示していない。コンピュータプログラムは、他のハードウェアと一緒に又は他のハードウェアの一部として供給された光記憶媒体又は固体媒体などの適切な媒体に格納/分配されてもよいが、さらに、インターネット又は他の有線若しくは無線通信システムなどを介して他の形態で分

50



配されてもよい。特許請求の範囲における参照符号は、本発明の範囲を限定するものとして解釈されるべきでない。

【図面】

【図 1】

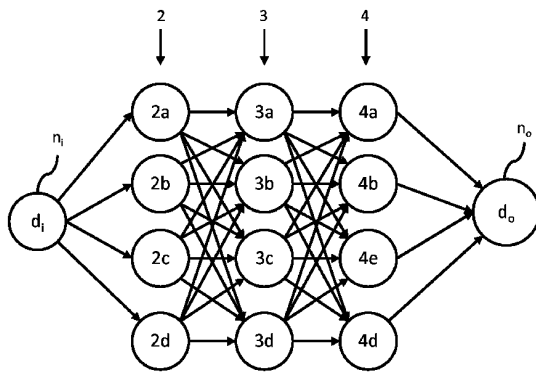


FIG. 1

【図 2】

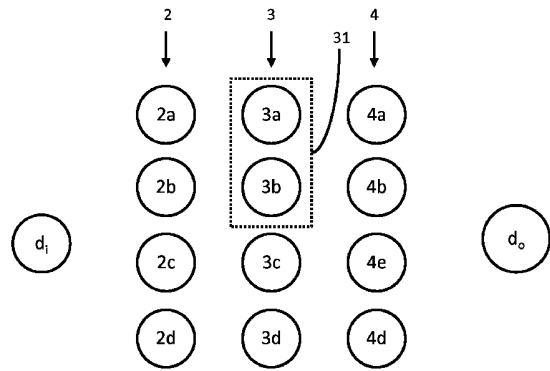


FIG. 2

【図 3】

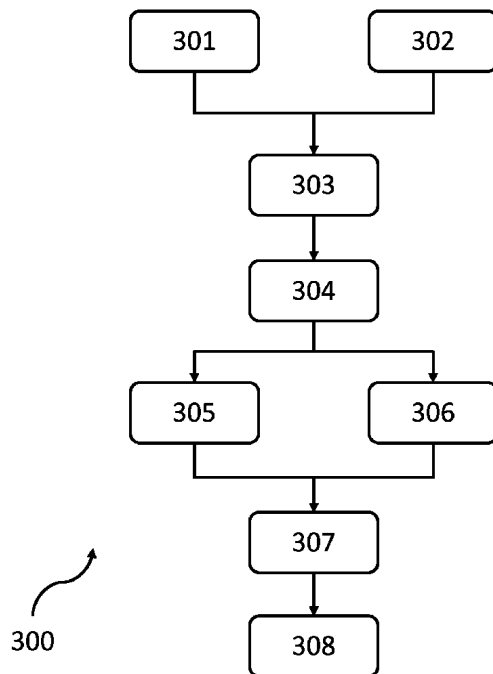


FIG. 3

【図 4】

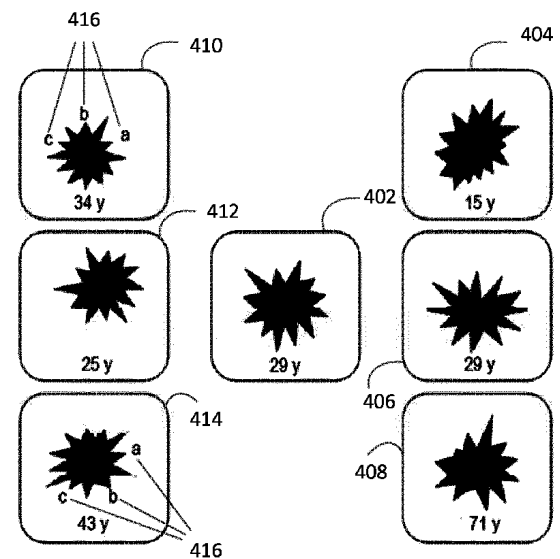


FIG. 4

10

20

30

40

50

【図 5】

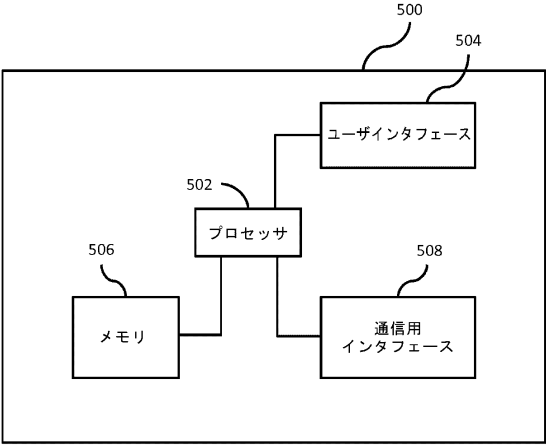


図 5

10

20

30

40

50

## フロントページの続き

(72)発明者 ラーマン アシフ

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

(72)発明者 レイノー カロライン デニス フランソワーズ

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

(72)発明者 コンロイ ブライアン

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

(72)発明者 マヴリュードゥス ディミトロス

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

(72)発明者 ブレシュ エリック

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

(72)発明者 ファン デン フーベル テウン

オランダ国 5 6 5 6 アーエー アインドーフェン ハイ テック キャンパス 5

審査官 今城 朋彬

(56)参考文献 特開平 0 4 - 1 9 5 2 4 9 ( J P , A )

特開 2 0 0 1 - 0 5 1 9 6 9 ( J P , A )

特開平 0 5 - 1 6 5 9 7 1 ( J P , A )

米国特許第 0 5 2 3 9 5 9 4 ( U S , A )

RAUBER, Paulo E., et al. , Visualizing the Hidden Activity of Artificial Neural Networks , IEE  
E Transactions on Visualization and Computer Graphics , Volume: 23, Issue: 1 , 2017年01  
月 , pp.101-110 , [online] [検索日 : 2022.11.22] URL: <https://ieeexplore.ieee.org/abstract/document/7539329>

(58)調査した分野 (Int.Cl. , D B 名)

G 0 6 N 3 / 0 0 - 9 9 / 0 0

G 0 6 T 7 / 0 0