

(19) 日本国特許庁(JP)

## (12) 公表特許公報(A)

(11) 特許出願公表番号

特表2017-520825

(P2017-520825A)

(43) 公表日 平成29年7月27日(2017.7.27)

(51) Int.Cl.

G06N 3/04 (2006.01)  
G06F 17/30 (2006.01)

F 1

G06N 3/04  
G06F 17/30

テーマコード(参考)

190  
210D

審査請求 未請求 予備審査請求 有 (全 53 頁)

(21) 出願番号 特願2016-567422 (P2016-567422)  
 (86) (22) 出願日 平成27年4月20日 (2015.4.20)  
 (85) 翻訳文提出日 平成29年1月10日 (2017.1.10)  
 (86) 國際出願番号 PCT/US2015/026621  
 (87) 國際公開番号 WO2015/175156  
 (87) 國際公開日 平成27年11月19日 (2015.11.19)  
 (31) 優先権主張番号 61/992,168  
 (32) 優先日 平成26年5月12日 (2014.5.12)  
 (33) 優先権主張国 米国(US)  
 (31) 優先権主張番号 14/483,054  
 (32) 優先日 平成26年9月10日 (2014.9.10)  
 (33) 優先権主張国 米国(US)

(71) 出願人 595020643  
 クアアルコム・インコーポレイテッド  
 QUALCOMM INCORPORATED  
 アメリカ合衆国、カリフォルニア州 92  
 121-1714、サン・ディエゴ、モア  
 ハウス・ドライブ 5775  
 (74) 代理人 100108855  
 弁理士 蔡田 昌俊  
 (74) 代理人 100109830  
 弁理士 福原 淑弘  
 (74) 代理人 100158805  
 弁理士 井関 守三  
 (74) 代理人 100112807  
 弁理士 岡田 貴志

最終頁に続く

(54) 【発明の名称】共通特徴にわたるカスタマイズされた識別子

## (57) 【要約】

分類子モデルを生成する方法は、共通特徴モデルを2人以上のユーザに分散することを含む。複数の分類子は、共通特徴モデル上でトレーニングされる。本方法は、複数の分類子のうちの第1の分類子を第1のユーザに、また複数の分類子のうちの第2の分類子を第2のユーザに分散することをさらに含む。

【選択図】図16

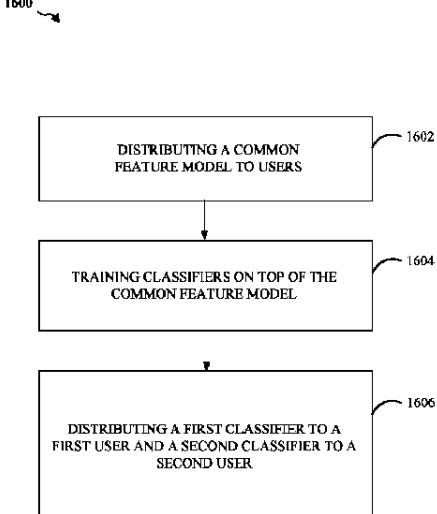


FIG. 16

**【特許請求の範囲】****【請求項 1】**

分類子モデルを生成する方法であって、  
共通特徴モデルを複数のユーザに分散することと、  
前記共通特徴モデル上で複数の分類子をトレーニングすることと、  
前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに、  
および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザ  
に分散することと  
を備える、方法。

**【請求項 2】**

前記複数の分類子のうちの少なくとも1つは、少なくとも1つのエンティティから取得  
されたラベル付けされたデータのセット上でトレーニングされる、

請求項1に記載の方法。

10

**【請求項 3】**

前記少なくとも1つのエンティティは、前記複数のユーザのうちの指定されたユーザを  
備える、

請求項2に記載の方法。

**【請求項 4】**

前記少なくとも1つのエンティティから前記取得されたラベル付けされたデータのセッ  
ト上で前記トレーニングされた分類子の各々にメトリックを提供することをさらに備える

20

請求項2に記載の方法。

**【請求項 5】**

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える

請求項4に記載の方法。

**【請求項 6】**

遠隔デバイス上のデータから計算された少なくとも1つの特徴を受信することと、  
前記少なくとも1つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記  
データを分類するための少なくとも1つの分類子を決定することと、

30

少なくとも1つの分類子の表示を前記遠隔デバイスに分散することと  
をさらに備える、請求項1に記載の方法。

**【請求項 7】**

遠隔デバイス上のデータから計算された特徴を受信することと、  
前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、

前記分類を前記遠隔デバイスに送信することと

をさらに備える、請求項1に記載の方法。

**【請求項 8】**

ラベル付けされたデータの1つまたは複数のセットに関連付けられるクラスの分類を行  
うように構成された、組み合わされた分類子を生成するために、前記第1の分類子と前記  
第2の分類子を組み合わせることと、

40

前記組み合わされた分類子を前記複数のユーザのうちの少なくとも1人に分散すること  
と

をさらに備える、請求項1に記載の方法。

**【請求項 9】**

前記複数の分類子のうちの少なくとも1つは、少なくとも1つのエンティティからのフ  
ィードバックに少なくとも部分的に基づいてトレーニングされる、

請求項1に記載の方法。

**【請求項 10】**

分類子モデルを生成するための装置であって、

50

メモリと、

前記メモリに結合された少なくとも1つのプロセッサと  
を備え、前記少なくとも1つのプロセッサは、

共通特徴モデルを複数のユーザに分散することと、

前記共通特徴モデル上で複数の分類子をトレーニングすることと、

前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに  
、および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザに  
分散することと

を行うように構成される、装置。

**【請求項11】**

前記少なくとも1つのプロセッサは、前記複数の分類子のうちの少なくとも1つを、少  
なくとも1つのエンティティから取得されたラベル付けされたデータのセット上でトレ  
ーニングすることを行うようにさらに構成される、

請求項10に記載の装置。

**【請求項12】**

前記少なくとも1つのエンティティは、前記複数のユーザのうちの指定されたユーザを  
備える、

請求項11に記載の装置。

**【請求項13】**

前記少なくとも1つのプロセッサは、前記少なくとも1つのエンティティから前記取得  
されたラベル付けされたデータのセット上で前記トレーニングされた分類子の各々にメト  
リックを提供することを行うようにさらに構成される、

請求項11に記載の装置。

**【請求項14】**

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える

、  
請求項13に記載の装置。

**【請求項15】**

前記少なくとも1つのプロセッサは、  
遠隔デバイス上のデータから計算された少なくとも1つの特徴を受信することと、

前記少なくとも1つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記  
データを分類するための少なくとも1つの分類子を決定することと、

少なくとも1つの分類子の表示を前記遠隔デバイスに分散することと

を行うようにさらに構成される、請求項10に記載の装置。

**【請求項16】**

前記少なくとも1つのプロセッサは、  
遠隔デバイス上のデータから計算された特徴を受信することと、  
前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、  
前記分類を前記遠隔デバイスに送信することと  
ようさらに構成される、請求項10に記載の装置。

**【請求項17】**

前記少なくとも1つのプロセッサは、  
ラベル付けされたデータの1つまたは複数のセットに関連付けられるクラスの分類を行  
うように構成された、組み合わされた分類子を生成するために、前記第1の分類子と前記  
第2の分類子を組み合わせることと、

前記組み合わされた分類子を前記複数のユーザのうちの少なくとも1人に分散すること  
と

を行うようにさらに構成される、請求項10に記載の装置。

**【請求項18】**

前記少なくとも1つのプロセッサは、少なくとも1つのエンティティからのフィードバ

10

20

30

40

50

ックに少なくとも部分的に基づいて、前記複数の分類子のうちの少なくとも1つをトレーニングすることを行うようにさらに構成される、

請求項10に記載の装置。

【請求項19】

分類子モデルを生成するための装置であって、  
共通特徴モデルを複数のユーザに分散するための手段と、  
前記共通特徴モデル上で複数の分類子をトレーニングするための手段と、  
前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに、  
および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザ  
に分散するための手段と  
を備える、装置。

10

【請求項20】

分類子モデルを生成するためのコンピュータプログラム製品であって、  
プログラムコードを符号化した非一時的コンピュータ可読媒体を備え、前記プログラム  
コードは、  
共通特徴モデルを複数のユーザに分散するためのプログラムコードと、  
前記共通特徴モデル上で複数の分類子をトレーニングするためのプログラムコードと、  
前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに、  
および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザ  
に分散するためのプログラムコードと  
を備える、コンピュータプログラム製品。

20

【発明の詳細な説明】

【関連出願の相互参照】

【0001】

[0001]本出願は、2014年5月12日に出願された「CUSTOMEZED CLASSIFIER OVER COMMON FEATURES」と題する米国仮特許出願第61/992,168号の利益を主張し、その開示は、参照によりその全体が本明細書に明示的に組み込まれる。

【技術分野】

【0002】

[0002]本開示のいくつかの態様は、一般にニューラルシステムエンジニアリングに関し、より詳細には、共通特徴のセットにわたるカスタマイズされた分類子を生成するためのシステムおよび方法に関する。

30

【背景技術】

【0003】

[0003]人工ニューロン（すなわち、ニューロンモデル）の相互結合されたグループを備え得る人工ニューラルネットワークは、計算デバイスであるか、または計算デバイスによって実行される方法を表す。人工ニューラルネットワークは、生物学的ニューラルネットワークにおける対応する構造および/または機能を有し得る。しかしながら、人工ニューラルネットワークは、従来の計算技法が厄介、実行不可能または不適切であるいくつかの適用例に革新的で有用な計算技法を提供することができる。人工ニューラルネットワークは観測から関数を推測することができるので、そのようなネットワークは、タスクまたはデータの複雑さが従来の技法による関数の設計を煩わしくする用途において、特に有用である。

40

【発明の概要】

【0004】

[0004]本開示の一態様では、分類子モデルを生成する方法が開示される。本方法は、共通特徴モデルを2人以上のユーザに分散することを含む。本方法はまた、共通特徴モデル上で複数の分類子をトレーニングすることを含む。本方法は、複数の分類子のうちの第1の分類子を2人以上のユーザのうちの第1のユーザに、および複数の分類子のうちの第2

50

の分類子を2人以上のユーザのうちの第2のユーザに分散することをさらに含む。

【0005】

[0005]本開示の別の態様では、分類子モデルを生成するための装置が開示される。本装置は、メモリと、メモリに結合された1つまたは複数のプロセッサとを含む。本プロセッサは、共通特徴モデルを2人以上のユーザに分散するように構成される。本プロセッサはまた、共通特徴モデル上で複数の分類子をトレーニングするように構成される。本プロセッサは、複数の分類子のうちの第1の分類子を2人以上のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を2人以上のユーザのうちの第2のユーザに分散するようにさらに構成される。

【0006】

[0006]本開示の別の態様では、分類子モデルを生成するための装置が開示される。本方法は、共通特徴モデルを2人以上のユーザに分散するための手段を含む。本方法はまた、共通特徴モデル上で複数の分類子をトレーニングするための手段を含む。本方法は、複数の分類子のうちの第1の分類子を2人以上のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を2人以上のユーザのうちの第2のユーザに分散するための手段をさらに含む。

【0007】

[0007]本開示の別の態様では、分類子モデルを生成するためのコンピュータプログラム製品が開示される。本コンピュータプログラム製品は、プログラムコードを符号化した非一時的コンピュータ可読媒体を含む。本プログラムコードは、共通特徴モデルを2人以上のユーザに分散するためのプログラムコードを含む。本プログラムコードはまた、共通特徴モデル上で複数の分類子をトレーニングするためのプログラムコードを含む。本プログラムコードは、複数の分類子のうちの第1の分類子を2人以上のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を2人以上のユーザのうちの第2のユーザに分散するプログラムコードをさらに含む。

【0008】

[0008]これは、以下の詳細な説明がより良く理解され得るために、本開示の特徴および技術的利点をかなり広く概説した。本開示の追加の特徴および利点は、以下で説明される。この開示は、本開示と同じ目的を実行するための他の構造を修正または設計するための基礎として容易に変更され得ることが、当業者によって理解されるべきである。また、添付の特許請求の範囲に記載されるように、そのような等価な構成は本開示の教示から逸脱しないことが、当業者によって理解されるべきである。本開示の特徴と考えられる新規な特徴は、その構成と動作の方法との両方に関して、さらなる目的および利点とともに、添付の図面と関連して考慮されるとき以下の説明からより良く理解されるであろう。しかしながら、図面の各々は単に例示および説明の目的のために提供されているにすぎず、本開示の制限の定義として意図されていないことが、明確には理解されるべきである。

【図面の簡単な説明】

【0009】

[0009]本開示の特徴、性質、および利点は、同様の参照文字が全体を通して相応して識別する図面を考慮した場合、以下に示される詳細な説明から、より明らかになるだろう。

【図1】本開示のいくつかの態様によるニューロンの例示的なネットワークを示す図。

【図2】本開示のいくつかの態様による、計算ネットワーク（ニューラルシステムまたはニューラルネットワーク）の処理ユニット（ニューロン）の一例を示す図。

【図3】本開示のいくつかの態様によるスパイクタイミング依存可塑性（STDP）曲線の一例を示す図。

【図4】本開示のいくつかの態様による、ニューロンモデルの挙動を定義するための正レジームおよび負レジームの一例を示す図。

【図5】本開示のある態様による、汎用プロセッサを使用してニューラルネットワークを設計することの例示的な実装形態を示す図。

【図6】本開示のいくつかの態様による、メモリが個々の分散処理ユニットとインターフ

10

20

30

40

50

エースされ得るニューラルネットワークを設計する例示的な実装形態を示す図。

【図7】本開示のいくつかの態様による、分散メモリおよび分散処理ユニットに基づいてニューラルネットワークを設計する例示的な実装形態を示す図。

【図8】本開示のいくつかの態様による、ニューラルネットワークの例示的な実装形態を示す図。

【図9】本開示の態様による、モデルを学習するための例示的なデータフローを示すブロック図。

【図10】本開示の態様による、分類子のための例示的なアーキテクチャを示すブロック図。

【図11】本開示の態様による、モデルを学習するための例示的なデータフローを示すブロック図。 10

【図12】本開示の態様による、分類子を生成するための例示的なデータフローを示すフローチャート。

【図13】本開示の態様による、モデルを学習するための方法を示す図。

【図14】本開示の態様による、モデルを学習するための方法を示す図。

【図15】本開示の態様による、分類子モデルを生成するための方法を示す図。

【図16】本開示の態様による、分類子モデルを生成するための方法を示す図。

【図17】本開示の態様による、分類子モデルを生成するための方法を示す図。

【発明を実施するための形態】

【0010】 20

[0027]添付の図面に関連して以下に示される詳細な説明は、様々な構成の説明として意図されたものであり、本明細書において説明される概念が実現され得る唯一の構成を表すことを意図されるものではない。詳細な説明は、様々な概念の完全な理解を提供する目的で、具体的な詳細を含む。しかしながら、これらの概念がこれらの具体的な詳細なしで実施され得ることは、当業者にとっては明らかであろう。いくつかの事例では、よく知られている構造および構成要素が、そのような概念を曖昧にするのを避けるために、ブロック図形式で示される。

【0011】

[0028]本教示に基づいて、本開示の範囲は、本開示の任意の他の態様とは無関係に実装されるにせよ、本開示の任意の他の態様と組み合わされるにせよ、本開示のいかなる態様をもカバーするものであることを、当業者なら諒解されたい。たとえば、記載される態様をいくつ使用しても、装置は実装され得、または方法は実施され得る。さらに、本開示の範囲は、記載される本開示の様々な態様に加えてまたはそれらの態様以外に、他の構造、機能、または構造および機能を使用して実施されるそのような装置または方法をカバーするものとする。開示する本開示のいずれの態様も、請求項の1つまたは複数の要素によって実施され得ることを理解されたい。 30

【0012】

[0029]「例示的」という単語は、本明細書では「例、事例、または例示の働きをすること」を意味するために使用される。「例示的」として本明細書で説明するいかなる態様も、必ずしも他の態様よりも好ましいまたは有利であると解釈されるべきであるとは限らない。 40

【0013】

[0030]本明細書では特定の態様について説明するが、これらの態様の多くの変形および置換は本開示の範囲内に入る。好ましい態様のいくつかの利益および利点が説明されるが、本開示の範囲は特定の利益、使用、または目的に限定されるものではない。むしろ、本開示の態様は、様々な技術、システム構成、ネットワーク、およびプロトコルに広く適用可能であるものとし、そのうちのいくつかを例として図および好ましい態様についての以下の説明で示す。発明を実施するための形態および図面は、本開示を限定するものではなく説明するものにすぎず、本開示の範囲は添付の特許請求の範囲およびそれの均等物によって定義される。 50

## 例示的なニューラルシステム、トレーニングおよび動作

### 【0014】

[0031] 図1は、本開示のいくつかの態様による、複数のレベルのニューロンをもつ例示的な人工ニューラルシステム100を示す。ニューラルシステム100は、シナプス結合のネットワーク104(すなわち、フィードフォワード結合)を介してニューロンの別のレベル106に結合されたニューロンのあるレベル102を有し得る。簡単のために、図1には2つのレベルのニューロンのみが示されているが、ニューラルシステムには、より少ないまたはより多くのレベルのニューロンが存在し得る。ニューロンのいくつかは、ラテラル結合を介して同じレイヤの他のニューロンに結合し得ることに留意されたい。さらに、ニューロンのいくつかは、フィードバック結合を介して前のレイヤのニューロンに戻る形で結合し得る。

10

### 【0015】

[0032] 図1に示すように、レベル102における各ニューロンは、前のレベル(図1に図示せず)のニューロンによって生成され得る入力信号108を受信し得る。信号108は、レベル102のニューロンの入力電流を表し得る。この電流は、膜電位を充電するためにニューロン膜上に蓄積され得る。膜電位がそれのしきい値に達すると、ニューロンは、発火し、ニューロンの次のレベル(たとえば、レベル106)に転送されるべき出力スパイクを生成し得る。いくつかのモデリング手法では、ニューロンは、信号をニューロンの次のレベルに継続的に転送し得る。この信号は、典型的には膜電位の関数である。そのような挙動は、以下で説明するものなどのアナログおよびデジタル実装形態を含むハードウェアおよび/またはソフトウェアでエミュレートまたはシミュレートされ得る。

20

### 【0016】

[0033] 生物学的ニューロンでは、ニューロンが発火するときに生成される出力スパイクは、活動電位と呼ばれる。電気信号は、約100mVの振幅と約1msの持続時間とを有する比較的急速で、一時的な神経インパルスである。一連の結合されたニューロンを有するニューラルシステムの特定の実施形態(たとえば、図1におけるあるレベルのニューロンから別のレベルのニューロンへのスパイクの転送)では、あらゆる活動電位が基本的に同じ振幅と持続時間とを有するので、信号における情報は、振幅によってではなく、スパイクの周波数および数、またはスパイクの時間によってのみ表され得る。活動電位によって搬送される情報は、スパイク、スパイクしたニューロン、および他の1つまたは複数のスパイクに対するスパイクの時間によって決定され得る。以下で説明するように、スパイクの重要性は、ニューロン間の接続に適用される重みによって決定され得る。

30

### 【0017】

[0034] 図1に示されるように、ニューロンのあるレベルから別のレベルへのスパイクの移動は、シナプス結合(または、単純に「シナプス」)104のネットワークを介して達成され得る。シナプス104に関して、レベル102のニューロンはシナプス前ニューロンと考えられ得、レベル106のニューロンはシナプス後ニューロンと考えられ得る。シナプス104は、レベル102のニューロンから出力信号(すなわち、スパイク)を受信して、調整可能なシナプスの重み

40

### 【0018】

### 【数1】

$$w_1^{(i,i+1)}, \dots, w_P^{(i,i+1)}$$

### 【0019】

に応じてそれらの信号をスケーリングすることができ、上式で、Pはレベル102のニューロンとレベル106のニューロンとの間のシナプス結合の総数であり、iはニューロンレベルの指標である。図1の例では、iはニューロンレベル102を表し、i+1は、ニューロンレベル106を表す。さらに、スケーリングされた信号は、レベル106における各ニューロンの入力信号として合成され得る。レベル106におけるあらゆるニューロ

50

ンは、対応する合成された入力信号に基づいて、出力スパイク 110 を生成し得る。出力スパイク 110 は、シナプス結合の別のネットワーク（図 1 には図示せず）を使用して、別のレベルのニューロンに転送され得る。

#### 【0020】

[0035]生物学的シナプスは、シナプス後ニューロンにおける興奮性活動または抑制性（過分極化）活動のいずれかを調停することができ、ニューロン信号を増幅する役目を果たすことができる。興奮性信号は、膜電位を脱分極する（すなわち、静止電位に対して膜電位を増加させる）。しきい値を超えて膜電位を脱分極するために十分な興奮性信号が一定の時間期間内に受信された場合、シナプス後ニューロンに活動電位が生じる。対照的に、抑制性信号は一般に、膜電位を過分極する（すなわち、低下させる）。抑制性信号は、十分に強い場合、興奮性信号のすべてを相殺し、膜電位がしきい値に達するのを防止することができる。シナプス興奮を相殺することに加えて、シナプス抑制は、自然に活発なニューロンに対して強力な制御を行うことができる。自然に活発なニューロンは、たとえば、そのダイナミクスまたはフィードバックに起因するさらなる入力なしにスパイクするニューロンを指す。これらのニューロンにおける活動電位の自然な生成を抑圧することによって、シナプス抑制は、一般にスカルプチャリングと呼ばれる、ニューロンの発火のパターンを形成することができる。様々なシナプス 104 は、望まれる拳動に応じて、興奮性シナプスまたは抑制性シナプスの任意の組合せとして働き得る。

10

#### 【0021】

[0036]ニューラルシステム 100 は、汎用プロセッサ、デジタル信号プロセッサ（D S P）、特定用途向け集積回路（A S I C）、フィールドプログラマブルゲートアレイ（F P G A）もしくは他のプログラマブル論理デバイス（P L D）、個別ゲートもしくはトランジスタ論理、個別ハードウェア構成要素、プロセッサによって実行されるソフトウェアモジュール、またはそれらの任意の組合せによってエミュレートされ得る。ニューラルシステム 100 は、たとえば画像およびパターン認識、機械学習、モータ制御、および同様のものなど、かなりの適用範囲において利用され得る。ニューラルシステム 100 における各ニューロンは、ニューロン回路として実装され得る。出力スパイクを開始するしきい値まで充電されるニューロン膜は、たとえば、そこを通って流れる電流を積分するキャパシタとして実装され得る。

20

#### 【0022】

[0037]一態様では、キャパシタは、ニューロン回路の電流積分デバイスとして除去され得、その代わりにより小さいメモリスタ（memristor）要素が使用され得る。この手法は、ニューロン回路において、ならびにかさばるキャパシタが電流積分器として利用される様々な他の適用例において適用され得る。さらに、シナプス 104 の各々は、メモリスタ要素に基づいて実装され得、シナプス重みの変化は、メモリスタ抵抗の変化に関係し得る。ナノメートルの特徴サイズのメモリスタを用いると、ニューロン回路およびシナプスの面積が大幅に低減され得、それによって、大規模なニューラルシステムハードウェア実装形態の実装がより実用的になり得る。

30

#### 【0023】

[0038]ニューラルシステム 100 をエミュレートするニューラルプロセッサの機能は、ニューロン間の結合の強さを制御し得る、シナプス結合の重みに依存し得る。シナプス重みは、パワーダウン後にプロセッサの機能を維持するために、不揮発性メモリに記憶され得る。一態様では、シナプス重みメモリは、主たるニューラルプロセッサチップとは別個の外部チップ上に実装され得る。シナプス重みメモリは、交換可能メモリカードとしてニューラルプロセッサチップとは別個にパッケージ化され得る。これは、ニューラルプロセッサに多様な機能を提供することができ、特定の機能は、ニューラルプロセッサに現在取り付けられているメモリカードに記憶されたシナプス重みに基づき得る。

40

#### 【0024】

[0039]図 2 は、本開示のいくつかの態様による、計算ネットワーク（たとえば、ニューラルシステムまたはニューラルネットワーク）の処理ユニット（たとえば、ニューロンま

50

たはニューロン回路) 202 の例示的な図 200 を示す。たとえば、ニューロン 202 は、図 1 のレベル 102 のニューロンおよび 106 のニューロンのうちのいずれかに対応し得る。ニューロン 202 は、ニューラルシステムの外部にある信号、または同じニューラルシステムの他のニューロンによって生成された信号、またはその両方であり得る、複数の入力信号 204<sub>1</sub> ~ 204<sub>N</sub> を受信し得る。入力信号は、電流、コンダクタンス、電圧、実数値および / または複素数値であり得る。入力信号は、固定小数点表現または浮動小数点表現をもつ数値を備え得る。これらの入力信号は、調整可能なシナプス重み 206<sub>1</sub> ~ 206<sub>N</sub> ( $w_1 \sim w_N$ ) に従って信号をスケーリングするシナプス結合を通してニューロン 202 に伝えられ得、N はニューロン 202 の入力接続の総数であり得る。

## 【0025】

10

[0040] ニューロン 202 は、スケーリングされた入力信号を合成し、合成された、スケーリングされた入力を使用して、出力信号 208 ( すなわち、信号 y ) を生成し得る。出力信号 208 は、電流、コンダクタンス、電圧、実数値および / または複素数値であり得る。出力信号は、固定小数点表現または浮動小数点表現をもつ数値であり得る。出力信号 208 は、次いで、同じニューラルシステムの他のニューロンへの入力信号として、または同じニューロン 202 への入力信号として、またはニューラルシステムの出力として伝達され得る。

## 【0026】

20

[0041] 処理ユニット ( ニューロン ) 202 は電気回路によってエミュレートされ得、その入力接続および出力接続は、シナプス回路をもつ電気接続によってエミュレートされ得る。処理ユニット 202 ならびにその入力接続および出力接続はまた、ソフトウェアコードによってエミュレートされ得る。処理ユニット 202 はまた、電気回路によってエミュレートされ得るが、その入力接続および出力接続はソフトウェアコードによってエミュレートされ得る。一態様では、計算ネットワーク中の処理ユニット 202 はアナログ電気回路であり得る。別の態様では、処理ユニット 202 はデジタル電気回路であり得る。さらに別の態様では、処理ユニット 202 は、アナログ構成要素とデジタル構成要素の両方をもつ混合信号電気回路であり得る。計算ネットワークは、上述の形態のいずれかにおける処理ユニットを含み得る。そのような処理ユニットを使用した計算ネットワーク ( ニューラルシステムまたはニューラルネットワーク ) は、たとえば画像およびパターン認識、機械学習、モータ制御など、かなりの適用範囲において利用され得る。

30

## 【0027】

[0042] ニューラルネットワークをトレーニングする過程で、シナプス重み ( たとえば、図 1 の重み )

## 【0028】

## 【数 2】

$$w_1^{(i,i+1)}, \dots, w_P^{(i,i+1)}$$

## 【0029】

40

および / または図 2 の重み 206<sub>1</sub> ~ 206<sub>N</sub> ) がランダム値により初期化され得、学習ルールに従って増加または減少し得る。学習ルールの例は、これに限定されないが、スパイクタイミング依存可塑性 ( STDP ) 学習ルール、Hebb 則、Oja 則、Biennenstock - Cooper - Munro ( BCM ) 則等を含むことを当業者は理解するだろう。いくつかの態様では、重みは、2 つの値のうちの 1 つに安定または収束し得る ( すなわち、重みの双峰分布 )。この効果が利用されて、シナプス重みごとのビット数を低減し、シナプス重みを記憶するメモリとの間の読み取りおよび書き込みの速度を上げ、シナプスメモリの電力および / またはプロセッサ消費量を低減し得る。

シナプスタイプ

## 【0030】

50

[0043] ニューラルネットワークのハードウェアおよびソフトウェアモデルでは、シナプ

ス関係機能の処理がシナプスタイプに基づき得る。シナプスタイプは、非塑性シナプス（non-plastic synapse）（重みおよび遅延の変化がない）、可塑性シナプス（重みが変化し得る）、構造遅延可塑性シナプス（重みおよび遅延が変化し得る）、完全可塑性シナプス（重み、遅延および結合性が変化し得る）、およびそれの変形（たとえば、遅延は変化し得るが、重みまたは結合性の変化はない）であり得る。複数のタイプの利点は、処理が再分割され得ることである。たとえば、非塑性シナプスは、可塑性機能を使用しないで実行される場合がある（またはそのような機能が完了するのを待つ）。同様に、遅延および重み可塑性は、一緒にまたは別々に、順にまたは並列に動作し得る動作に再分割され得る。異なるタイプのシナプスは、適用される異なる可塑性タイプの各々の異なるルックアップテーブルまたは式およびパラメータを有し得る。したがって、本方法は、シナプスのタイプについての関連する表、式、またはパラメータにアクセスする。

10

## 【0031】

[0044]スパイクタイミング依存構造可塑性がシナプス可塑性とは無関係に実行され得るという事実のさらなる含意がある。構造可塑性は、重みの大きさに変化がない場合（たとえば、重みが最小値または最大値に達したか、あるいはそれが何らかの他の理由により変更されない場合）<sup>10</sup>構造可塑性（すなわち、遅延量の変化）は前後スパイク時間差（pre-post spike time difference）の直接関数であり得ても実行され得る。代替的に、構造可塑性は、重み変化量に応じて、または重みもしくは重み変化の限界に関係する条件に基づいて設定され得る。たとえば、重み変化が生じたとき、または重みが最大値になるのではなく、重みがゼロに達した場合のみ、シナプス遅延が変化し得る。しかしながら、これらのプロセスが並列化され、メモリアクセスの数および重複を低減し得るように、独立した機能を有することが有利であり得る。

20

## シナプス可塑性の決定

## 【0032】

[0045]神経可塑性（または単に「可塑性」）は、脳内のニューロンおよびニューラルネットワークがそれらのシナプス結合と挙動とを新しい情報、感覚上の刺激、発展、損傷または機能不全に応答して変える能力である。可塑性は、生物学における学習および記憶にとって、また計算論的神経科学およびニューラルネットワークにとって重要である。（たとえば、Hebb則理論による）シナプス可塑性、スパイクタイミング依存可塑性（STDP）、非シナプス可塑性、活性依存可塑性、構造可塑性および恒常的可塑性など、様々な形の可塑性が研究されている。<sup>30</sup>

30

## 【0033】

[0046]STDPは、ニューロン間のシナプス結合の強さを調整する学習プロセスである。結合強度は、特定のニューロンの出力スパイクおよび受信入力スパイク（すなわち、活動電位）の相対的タイミングに基づいて調整される。STDPプロセスの下で、あるニューロンに対する入力スパイクが、平均して、そのニューロンの出力スパイクの直前に生じる傾向がある場合、長期増強（LTP）が生じ得る。その場合、その特定の入力はいくらか強くなる。一方、入力スパイクが、平均して、出力スパイクの直後に生じる傾向がある場合、長期抑圧（ LTD）が生じ得る。その場合、その特定の入力はいくらか弱くなるので、「スパイクタイミング依存可塑性」と呼ばれる。したがって、シナプス後ニューロンの興奮の原因であり得る入力は、将来的に寄与する可能性がさらに高くなる一方、シナプス後スパイクの原因ではない入力は、将来的に寄与する可能性が低くなる。結合の初期セットのサブセットが残る一方で、その他の部分の影響がわずかなレベルまで低減されるまで、このプロセスは続く。<sup>40</sup>

40

## 【0034】

[0047]ニューロンは出力スパイクを、その入力の多くが短い期間内に生じる（すなわち、出力をもたらすのに十分な累積がある）ときに生成し得るので、通常残っている入力のサブセットは、時間的に相關する傾向のあった入力を含む。さらに、出力スパイクの前に生じる入力は強化されるので、最も早い十分に累積的な相關指示を提供する入力は結局、ニューロンへの最終入力となる。<sup>50</sup>

50

## 【0035】

[0048] S T D P 学習ルールは、シナプス前ニューロンのスパイク時間  $t_{pre}$  とシナプス後ニューロンのスパイク時間  $t_{post}$  との間の時間差（すなわち、 $t = t_{post} - t_{pre}$ ）に応じて、シナプス前ニューロンをシナプス後ニューロンに結合するシナプスのシナプス重みを効果的に適合させ得る。S T D P の通常の公式化は、時間差が正である（シナプス前ニューロンがシナプス後ニューロンの前に発火する）場合にシナプス重みを増加させ（すなわち、シナプスを増強し）、時間差が負である（シナプス後ニューロンがシナプス前ニューロンの前に発火する）場合にシナプス重みを減少させる（すなわち、シナプスを抑制する）ことである。

## 【0036】

10

[0049] S T D P プロセスでは、経時的なシナプス重みの変化は通常、以下の式によって与えられるように、指數関数的減衰を使用して達成され得る。

## 【0037】

## 【数3】

$$\Delta w(t) = \begin{cases} a_+ e^{-t/k_+} + \mu, & t > 0 \\ a_- e^{t/k_-}, & t < 0 \end{cases}, \quad (1)$$

## 【0038】

20

ここで、 $k_+$  および  $k_-$  はそれぞれ、正の時間差および負の時間差の時間定数であり、 $a_+$  および  $a_-$  は対応するスケーリングの大きさであり、 $\mu$  は正の時間差および / または負の時間差に適用され得るオフセットである。

## 【0039】

30

[0050] 図3は、S T D P による、シナプス前スパイクおよびシナプス後スパイクの相対的タイミングに応じたシナプス重み変化の例示的な図300を示す。シナプス前ニューロンがシナプス後ニューロンの前に発火する場合、グラフ300の部分302に示すように、対応するシナプス重みは増加し得る。この重み増加は、シナプスのLTPと呼ばれ得る。グラフ部分302から、シナプス前スパイク時間とシナプス後スパイク時間との間の時間差に応じて、LTPの量がほぼ指數関数的に減少し得ることが観測され得る。グラフ300の部分304に示すように、発火の逆の順序は、シナプス重みを減少させ、シナプスのLTDをもたらし得る。

## 【0040】

40

[0051] 図3のグラフ300に示すように、S T D P グラフのLTP（原因）部分302に負のオフセット  $\mu$  が適用され得る。x軸の交差306のポイント（ $y = 0$ ）は、レイヤ*i* - 1からの原因入力の相關を考慮して、最大タイムラグと一致するように構成され得る。フレームベースの入力（すなわち、スパイクまたはパルスを備える特定の持続時間のフレームの形態である入力）の場合、オフセット値  $\mu$  は、フレーム境界を反映するように計算され得る。直接的にシナプス後電位によってモデル化されるように、またはニューラル状態に対する影響の点で、フレームにおける第1の入力スパイク（パルス）が経時的に減衰することが考慮され得る。フレームにおける第2の入力スパイク（パルス）が特定の時間フレームと相關したまたは特定の時間フレームに関連したものと考えられる場合、フレームの前および後の関連する時間は、その時間フレーム境界で分離され、関連する時間の値が異なり得る（たとえば、1つのフレームよりも大きい場合は負、1つのフレームよりも小さい場合は正）ように、S T D P 曲線の1つまたは複数の部分をオフセットすることによって、可塑性の点で別様に扱われ得る。たとえば、曲線が、フレーム時間よりも大きい前後の時間で実際にゼロよりも下になり、結果的にLTPの代わりにLTDの一部であるようにLTPをオフセットするために負のオフセット  $\mu$  が設定され得る。

ニューロンモデルおよび演算

## 【0041】

50

[0052] 有用なスパイキングニューロンモデルを設計するための一般的原理がいくつかあ

る。良いニューロンモデルは、2つの計算レジーム、すなわち、一致検出および関数計算の点で豊かな潜在的挙動を有し得る。その上、良いニューロンモデルは、時間コーディングを可能にするための2つの要素を有する必要がある：入力の到着時間は出力時間に影響を与え、一致検出は狭い時間ウィンドウを有し得る。最終的に、計算上魅力的であるために、良いニューロンモデルは、連続時間に閉形式解と、ニアアトラクター（near attractor）と鞍点とを含む安定した挙動とを有し得る。言い換えれば、有用なニューロンモデルは、実用的なニューロンモデルであり、豊かで、現実的で、生物学的に一貫した挙動をモデル化するために使用され得、神経回路のエンジニアリングとリバースエンジニアリングの両方を行うために使用され得るニューロンモデルである。

【0042】

10

[0053]ニューロンモデルは事象、たとえば入力の到着、出力スパイク、または内部的であるか外部的であるかを問わず他の事象に依存し得る。豊かな挙動レパートリーを実現するために、複雑な挙動を示すことができる状態機械が望まれ得る。入力寄与（ある場合）とは別個の事象の発生自体が状態機械に影響を与え、事象の後のダイナミクスを制限し得る場合、システムの将来の状態は、単なる状態および入力の関数ではなく、むしろ状態、事象および入力の関数である。

【0043】

[0054]一態様では、ニューロン  $n$  は、下記のダイナミクスによって決定される膜電圧  $v_n(t)$  によるスパイキングリーキー積分発火ニューロンとしてモデル化され得る。

【0044】

20

【数4】

$$\frac{dv_n(t)}{dt} = \alpha v_n(t) + \beta \sum_m w_{m,n} y_m(t - \Delta t_{m,n}), \quad (2)$$

【0045】

ここで  $\alpha$  および  $\beta$  は、パラメータであり、 $w_{m,n}$  は、シナプス前ニューロン  $m$  をシナプス後ニューロン  $n$  に結合するシナプスのシナプス重みであり、 $y_m(t)$  は、ニューロン  $n$  の細胞体に到着するまで  $t_{m,n}$  に従って樹状遅延または軸索遅延によって遅延し得るニューロン  $m$  のスパイキング出力である。

30

【0046】

[0055]シナプス後ニューロンへの十分な入力が達成された時間からシナプス後ニューロンが実際に発火する時間までの遅延があることに留意されたい。イジケヴィッチの単純モデルなど、動的スパイキングニューロンモデルでは、脱分極しきい値  $v_t$  とピークスパイク電圧  $v_{peak}$  との間に差がある場合、時間遅延が生じ得る。たとえば、単純モデルでは、電圧および復元のための1対の微分方程式、すなわち、

【0047】

40

【数5】

$$\frac{dv}{dt} = (k(v - v_t)(v - v_r) - u + I)/C, \quad (3)$$

$$\frac{du}{dt} = a(b(v - v_r) - u). \quad (4)$$

【0048】

によってニューロン細胞体ダイナミクス（neuron soma dynamics）が決定され得る。ここで  $v$  は膜電位であり、 $u$  は、膜復元変数であり、 $k$  は、膜電位  $v$  の時間スケールを記述するパラメータであり、 $a$  は、復元変数  $u$  の時間スケールを記述するパラメータであり、 $b$  は、膜電位  $v$  のしきい値下変動に対する復元変数  $u$  の感度を記述するパラメータであり、

50

$r$  は、膜静止電位であり、 $I$  は、シナプス電流であり、 $C$  は、膜のキャパシタンスである。このモデルによれば、ニューロンは  $v > v_{peak}$  のときにスパイクすると定義される。

### Hunzinger Cold モデル

#### 【0049】

[0056] Hunzinger Cold ニューロンモデルは、豊かな様々な神経挙動を再生し得る最小二重レジームスパイキング線形動的モデルである。モデルの1次元または2次元の線形ダイナミクスは2つのレジームを有することができ、時間定数（および結合）はレジームに依存し得る。しきい値下レジームでは、時間定数は、慣例により負であり、一般に生物学的に一貫した線形方式で静止状態に細胞を戻す役目を果たすリーキーチャネルダイナミクスを表す。しきい値上レジームにおける時間定数は、慣例により正であり、一般にスパイク生成のレイテンシを生じさせる一方でスパイク状態に細胞を駆り立てる反リーキーチャネルダイナミクスを反映する。  
10

#### 【0050】

[0057] 図4に示すように、モデル400のダイナミクスは2つの（またはそれよりも多くの）レジームに分割され得る。これらのレジームは、負のレジーム（negative regime）402（leaky-integrate-and-fire (LIF) ニューロンモデルと混同されないように、交換可能にLIFレジームとも呼ばれる）、および正のレジーム（positive regime）404（anti-leaky-integrate-and-fire (ALIF) ニューロンモデルと混同されないように、交換可能にALIFレジームとも呼ばれる）と呼ばれ得る。負レジーム402では、状態は将来の事象の時点における静止（..）の傾向がある。この負レジームでは、モデルは一般に、時間的入力検出特性と他のしきい値下挙動とを示す。正レジーム404では、状態はスパイキング事象（..）の傾向がある。この正レジームでは、モデルは、後続の入力事象に応じてスパイクにレイテンシを生じさせるなどの計算特性を示す。事象の点からのダイナミクスの公式化およびこれら2つのレジームへのダイナミクスの分離は、モデルの基本的特性である。  
20

#### 【0051】

[0058] 線形二重レジーム2次元ダイナミクス（状態  $v$  および  $u$  の場合）は、慣例により次のように定義され得る。  
30

#### 【0052】

#### 【数6】

$$\tau_p \frac{dv}{dt} = v + q_p \quad (5)$$

$$-\tau_u \frac{du}{dt} = u + r \quad (6)$$

#### 【0053】

ここで  $q$  および  $r$  は、結合のための線形変換変数である。  
40

#### 【0054】

[0059] シンボル  $\rho$  は、ダイナミクスレジームを示すためにここで使用され、特定のレジームの関係を論述または表現するときに、それぞれ負レジームおよび正レジームについて符号「-」または「+」にシンボル  $\rho$  を置き換える慣例がある。

#### 【0055】

[0060] モデル状態は、膜電位（電圧）  $v$  および復元電流  $u$  によって定義される。基本形態では、レジームは基本的にモデル状態によって決定される。正確で一般的な定義の微妙だが重要な側面があるが、差し当たり、モデルが、電圧  $v$  がしきい値（..）を上回る場合に正レジーム404にあり、そうでない場合に負レジーム402にあると考える。  
50

## 【0056】

[0061] レジーム依存時間定数は、負レジーム時間定数である  $\tau_-$  と正レジーム時間定数である  $\tau_+$  を含む。復元電流時間定数  $\tau_u$  は通常、レジームから独立している。便宜上、 $\tau_u$  と同様に、指數および  $\tau_+$  が一般に正となる正レジームの場合に、電圧発展 (voltage evolution) に関する同じ表現が使用され得るように、減衰を反映するために負の量として負レジーム時間定数  $\tau_-$  が一般に指定される。

## 【0057】

[0062] 2つの状態要素のダイナミクスは、事象において、ヌルクライインから状態をオフセットする変換によって結合され得、ここで変換変数は、

## 【0058】

## 【数7】

$$q_p = -\tau_p \beta u - v_p \quad (7)$$

$$r = \delta(v + \epsilon) \quad (8)$$

## 【0059】

であり、 $v$ 、 $u$ 、 $\epsilon$  および  $\delta$  はパラメータである。そのための2つの値は、2つのレジームのための参照電圧のベースである。パラメータ  $\tau_u$  は、負レジームのためのベース電圧であり、膜電位は一般に、負レジームにおいて  $v$  に減衰する。パラメータ  $\tau_+$  は、正レジームのためのベース電圧であり、膜電位は一般に、正レジームにおいて  $v$  から離れる傾向となる。

## 【0060】

[0063] および  $u$  のためのヌルクライインは、それぞれ変換変数  $q$  および  $r$  の負によって与えられる。パラメータ  $\tau_u$  は、 $u$  ヌルクライインの傾きを制御するスケール係数である。パラメータ  $\tau_-$  は通常、 $-v$  に等しく設定される。パラメータ  $\tau_+$  は、両方のレジームにおいて ヌルクライインの傾きを制御する抵抗値である。時間定数パラメータは、指數関数的減衰だけでなく、各レジームにおいて別個にヌルクライインの傾きを制御する。

## 【0061】

[0064] モデルは、電圧  $v$  が値  $v_s$  に達したときにスパイクするように定義され得る。続いて、状態は（スパイク事象と同じ1つのものであり得る）リセット事象でリセットされ得る。

## 【0062】

## 【数8】

$$v = \hat{v}_- \quad (9)$$

$$u = u + \Delta u \quad (10)$$

## 【0063】

ここで、

## 【0064】

## 【数9】

$$\hat{v}_-$$

## 【0065】

および  $u$  はパラメータである。リセット電圧

## 【0066】

## 【数10】

$$\hat{v}_-$$

## 【0067】

10

20

30

40

50

は通常、 $\tau_p$ にセットされる。

【0068】

[0065]瞬時結合の原理によって、状態について（また、単一の指數項による）だけではなく、特定の状態に到達するための時間についても、閉形式解が可能である。閉形式状態解は、次のとおりである。

【0069】

【数11】

$$v(t + \Delta t) = (v(t) + q_p) e^{-\frac{\Delta t}{\tau_p}} - q_p \quad (11)$$

$$u(t + \Delta t) = (u(t) + r) e^{-\frac{\Delta t}{\tau_u}} - r \quad (12)$$

【0070】

[0066]したがって、モデル状態は、入力（シナプス前スパイク）または出力（シナプス後スパイク）などの事象に伴ってのみ更新され得る。また、演算が（入力があるか、出力があるかを問わず）任意の特定の時間に実行され得る。

【0071】

[0067]その上、瞬時結合原理によって、反復的技法または数値解法（たとえば、オイラー数値解法）なしに、特定の状態に到達する時間が事前に決定され得るように、シナプス後スパイクの時間が予想され得る。前の電圧状態 $v_0$ を踏まえ、電圧状態 $v_f$ に到達するまでの時間遅延は、次の式によって与えられる。

【0072】

【数12】

$$\Delta t = \tau_p \log \frac{v_f + q_p}{v_0 + q_p} \quad (13)$$

【0073】

[0068]スパイクが、電圧状態 $v_s$ に到達する時間に生じると定義される場合、電圧 $v$ が所与の状態 $v_+$ にある時間から測定されたスパイクが生じるまでの時間量、または相対的遅延に関する閉形式解は、次のとおりである。

【0074】

【数13】

$$\Delta t_s = \begin{cases} \tau_+ \log \frac{v_s + q_+}{v + q_+} & \text{もし } v > \hat{v}_+ \text{ であるならば} \\ \infty & \text{それ以外} \end{cases} \quad (14)$$

【0075】

ここで、

【0076】

【数14】

$$\hat{v}_+$$

【0077】

は通常、パラメータ $\tau_+$ にセットされるが、他の変形も可能であり得る。

【0078】

[0069]モデルダイナミクスの上記の定義は、モデルが正レジームにあるか、それとも負

レジームにあるかに依存する。上述のように、結合およびレジームは、事象に伴って計算され得る。状態の伝搬のために、レジームおよび結合（変換）変数は、最後の（前の）事象の時間における状態に基づいて定義され得る。続いてスパイク出力時間を予想するために、レジームおよび結合変数は、次の（最新の）事象の時間における状態に基づいて定義され得る。

#### 【0079】

[0070] C o l d モデルの、適時にシミュレーション、エミュレーションまたはモデルを実行するいくつかの可能な実装形態がある。これは、たとえば、事象更新モード、ステップ事象更新モード、およびステップ更新モードを含む。事象更新は、（特定の瞬間ににおける）事象または「事象更新」に基づいて状態が更新される更新である。ステップ更新は、間隔（たとえば、1 m s）においてモデルが更新される更新である。これは必ずしも、反復的技法または数値解法を利用するとは限らない。また、事象がステップもしくはステップ間で生じる場合または「ステップ事象」更新によってモデルを更新するのみによって、ステップベースのシミュレータにおいて限られた時間分解能で事象ベースの実装形態が可能である。

10

#### 【0080】

[0071] 本開示は、スパイキングニューロン、ニューロンモデル、およびネットワークの様々な例を記載したが、本開示はそれらに限定されない。むしろ、非スパイキングニューロンおよびネットワークはまた、本開示のいくつかの態様を実現するために使用され得る。

20

共通特徴にわたる分散モデル学習およびカスタマイズされた分類子

#### 【0081】

[0072] 本開示の態様は、デバイスもまたモデルを使用している間に、分散され、緩やかに調整された方法でモデルの学習を継続するためのプロセスを対象とする。1つの例示的な態様では、ディープニューラルネットワーク（D N N）は、モバイルデバイス上の画像内のオブジェクト認識のために使用され得、モデルの改良を継続するために、モバイルデバイスは情報を中央サーバに返送する。説明を容易にするために、例示的なデータフローおよび他の説明が、画像およびオブジェクト認識に適用される。しかしながら、本開示はそれらに限定されず、代わりに任意の感覚様式が代替的または追加的に利用され得る。

30

#### 【0082】

[0073] 本開示のさらなる態様は、分類子モデルを生成することを対象とする。分類子モデルは、共通特徴セットにわたってカスタマイズされ得る。1つの例示的な態様では、中央サーバは、1つまたは複数のユーザデバイスからラベル付けされた例のコーパスを受信するように構成され得る。ユーザデバイスは、パーソナルコンピュータ（P C）、テレビ、ビデオゲームシステム、ラップトップ、タブレットP C、スマートフォンなどのモバイルデバイス、または他の携帯電子デバイスを備え得る。

30

#### 【0083】

[0074] サーバは、データセットに関連する統計的特徴のセットで構成され得る。いくつかの態様では、データセットは、たとえば、特定の感覚様式（画像、音、配向、位置など）に対応し得る。サーバは、ラベル付けされた例の受信したコーパス、および統計的特徴のセットに基づいて、分類子を生成し得る。

40

#### 【0084】

[0075] 図5は、本開示のいくつかの態様による、前述のモデルを学習すること、分類子モデルを生成すること、および／または汎用プロセッサ502を使用して分類子のセットを更新することの例示的な実装形態500を示している。変数（神経信号）、シナプス重み、計算ネットワーク（ニューラルネットワーク）に関連付けられるシステムパラメータ、遅延、周波数ビン情報パラメータ更新、外れ値情報、モデル更新、特徴情報、例および／またはラベル情報は、メモリブロック504で記憶され得、汎用プロセッサ502で実行される命令はプログラムメモリ506からロードされ得る。本開示のある態様では、汎用プロセッサ502にロードされる命令は、1人または複数のユーザからモデル更新を受

50

信し、前のモデルおよびモデル更新に基づいて更新されたモデルを計算し、ならびに／または、更新されたモデルに基づいて、更新されたモデルに関連するデータを1人または複数のユーザに送信するためのコードを備え得る。

#### 【0085】

[0076]本開示の別の態様では、汎用プロセッサ502にロードされた命令は、共有された推論モデルに基づいてサーバからデータを受信し、受信されたデータに基づいて1つまたは複数のモデルパラメータを含むモデルを生成し、モデルに基づいて推論を計算し、推論に基づいて1つまたは複数のモデルパラメータ更新を計算し、および／またはモデルパラメータ更新に基づいてデータをサーバに送信するためのコードを備え得る。

#### 【0086】

[0077]本開示の別の態様では、汎用プロセッサ502にロードされた命令は、分類子の第1のセットをデータの第1のセットに適用し、および／または、遠隔デバイスに、分類子の第1のセットの出力、あるいは分類子の第1のセットの適用の性能尺度のうちの1つまたは複数に基づいて分類子更新を要求するためのコードを備え得る。

#### 【0087】

[0078]本開示の別の態様では、汎用プロセッサ502にロードされた命令は、共通特徴モデルを複数のユーザに分散し、共通特徴モデル上の複数の分類子をトレーニングし、ならびに／または複数の分類子のうちの第1の分類子を複数のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を複数のユーザのうちの第2のユーザに分散するためのコードを備え得る。

#### 【0088】

[0079]図6は、本開示のいくつかの態様による、前述のモデルを学習すること、および／または分類子モデルを生成することの例示的な実装形態600を示しており、メモリ602は、相互接続ネットワーク604を介して、計算ネットワーク（ニューラルネットワーク）の個々の（分散された）処理ユニット（ニューラルプロセッサ）606とインターフェースされ得る。変数（神経信号）、シナプス重み、計算ネットワーク（ニューラルネットワーク）遅延に関連付けられるシステムパラメータ、周波数ビン情報パラメータ更新、外れ値情報、モデル更新、特徴情報、例および／またはラベル情報は、メモリ602で記憶され得、相互接続ネットワーク604の接続を介してメモリ602から各処理ユニット（ニューラルプロセッサ）606にロードされ得る。本開示のある態様では、処理ユニット606は、1人または複数のユーザからモデル更新を受信し、前のモデルおよびモデル更新に基づいて更新されたモデルを計算し、ならびに／または、更新されたモデルに基づいて、更新されたモデルに関連するデータを1人または複数のユーザに送信するように構成され得る。

#### 【0089】

[0080]本開示の別の態様では、処理ユニット606は、共有された推論モデルに基づいてサーバからデータを受信し、受信されたデータに基づいて1つまたは複数のモデルパラメータを含むモデルを生成し、モデルに基づいて推論を計算し、推論に基づいて1つまたは複数のモデルパラメータ更新を計算し、および／またはモデルパラメータ更新に基づいてデータをサーバに送信するように構成され得る。

#### 【0090】

[0081]本開示のさらに別の態様では、処理ユニット606は、分類子の第1のセットをデータの第1のセットに適用し、および／または、遠隔デバイスに、分類子の第1のセットの出力、あるいは分類子の第1のセットの適用の性能尺度のうちの1つまたは複数に基づいて分類子更新を要求するように構成され得る。

#### 【0091】

[0082]本開示のなおさらに別の態様では、処理ユニット606は、共通特徴モデルを複数のユーザに分散し、共通特徴モデル上の複数の分類子をトレーニングし、ならびに／または複数の分類子のうちの第1の分類子を複数のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を複数のユーザのうちの第2のユーザに分散するよう

10

20

30

40

50

に構成され得る。

【0092】

[0083]図7は、前述のモデルを学習すること、および／または分類子モデルを生成することの例示的な実装形態700を示している。図7に示されるように、1つのメモリバンク702は、計算ネットワーク（ニューラルネットワーク）の1つの処理ユニット704と直接インターフェースされ得る。各メモリバンク702は、変数（神経信号）、シナプス重み、および／または対応する処理ユニット（ニューラルプロセッサ）704遅延に関連付けられるシステムパラメータ、周波数ビン情報パラメータ更新、外れ値情報、モデル更新、特徴情報、例および／またはラベル情報を記憶し得る。本開示のある態様では、処理ユニット704は、1人または複数のユーザからモデル更新を受信し、前のモデルおよびモデル更新に基づいて更新されたモデルを計算し、ならびに／または、更新されたモデルに基づいて、更新されたモデルに関連するデータを1人または複数のユーザに送信するように構成され得る。  
10

【0093】

[0084]本開示のさらなる態様では、処理ユニット704は、共有された推論モデルに基づいてサーバからデータを受信し、受信されたデータに基づいて1つまたは複数のモデルパラメータを含むモデルを生成し、モデルに基づいて推論を計算し、推論に基づいて1つまたは複数のモデルパラメータ更新を計算し、および／またはモデルパラメータ更新に基づいてデータをサーバに送信するように構成され得る。

【0094】

[0085]本開示のさらに別の態様では、処理ユニット704は、分類子の第1のセットをデータの第1のセットに適用し、および／または、遠隔デバイスに、分類子の第1のセットの出力、あるいは分類子の第1のセットの適用の性能尺度のうちの1つまたは複数に基づいて分類子更新を要求するように構成され得る。

【0095】

[0086]本開示のなおさらに別の態様では、処理ユニット704は、共通特徴モデルを複数のユーザに分散し、共通特徴モデル上の複数の分類子をトレーニングし、ならびに／または複数の分類子のうちの第1の分類子を複数のユーザのうちの第1のユーザに、および複数の分類子のうちの第2の分類子を複数のユーザのうちの第2のユーザに分散するように構成され得る。  
30

【0096】

[0087]図8は、本開示のいくつかの態様による、ニューラルネットワーク800の例示的な実装形態を示す。図8に示すように、ニューラルネットワーク800は、本明細書に説明した方法の様々な動作を実行し得る複数のローカル処理ユニット802を有することができる。各ローカル処理ユニット802は、ニューラルネットワークのパラメータを記憶する、ローカルステートメモリ804およびローカルパラメタメモリ806を備え得る。また、ローカル処理ユニット802は、ローカルモデルプログラムを記憶するためのローカル（ニューロン）モデルプログラム（LMP）メモリ808、ローカル学習プログラムを記憶するためのローカル学習プログラム（LLP）メモリ810、およびローカル接続メモリ812を有し得る。さらに、図8に示されるように、各ローカル処理ユニット802は、ローカル処理ユニットのローカルメモリの構成を提供するための構成プロセッサユニット814と、またローカル処理ユニット802間のルーティングを提供するルーティング接続処理ユニット816とインターフェースされ得る。  
40

【0097】

[0088]一構成では、ニューロンモデルは、1人または複数のユーザからモデル更新を受信することと、前のモデルおよびモデル更新に基づいて更新されたモデルを計算することと、ならびに／または、更新されたモデルに基づいて、更新されたモデルに関連するデータを1人または複数のユーザに送信することとのために構成される。ニューロンモデルは、受信手段、計算手段、および送信手段を含む。一態様では、受信手段、計算手段、および／または送信手段は、記載される機能を実行するように構成された汎用プロセッサ50

2、プログラムメモリ 506、メモリブロック 504、メモリ 602、相互接続ネットワーク 604、処理ユニット 606、処理ユニット 704、ローカル処理ユニット 802、およびまたはルーティング接続処理ユニット 816 であり得る。

#### 【0098】

[0089]別の構成では、ニューロンモデルは、共有された推論モデルに基づいてサーバからデータを受信することと、モデルに基づいて推論を計算することと、推論に基づいて1つまたは複数のモデルパラメータ更新を計算することと、および／またはモデルパラメータ更新に基づいてデータをサーバに送信することとのために構成される。ニューロンモデルは、受信手段、計算手段、および送信手段を含む。一態様では、受信手段、推論を計算するための手段、モデルパラメータ更新を計算するための手段、および／または送信手段は、記載される機能を実行するように構成された汎用プロセッサ 502、プログラムメモリ 506、メモリブロック 504、メモリ 602、相互接続ネットワーク 604、処理ユニット 606、処理ユニット 704、ローカル処理ユニット 802、およびまたはルーティング接続処理ユニット 816 であり得る。

10

#### 【0099】

[0090]さらに別の構成では、ニューロンモデルは、分類子の第1のセットをデータの第1のセットに適用することと、および／または、遠隔デバイスに、分類子の第1のセットの出力、あるいは分類子の第1のセットの適用の性能尺度のうちの1つまたは複数に基づいて分類子更新を要求することとのために構成される。ニューロンモデルは適用手段と要求手段とを含む。一態様では、適用手段、および／または要求手段は、記載される機能を実行するように構成された汎用プロセッサ 502、プログラムメモリ 506、メモリブロック 504、メモリ 602、相互接続ネットワーク 604、処理ユニット 606、処理ユニット 704、ローカル処理ユニット 802、およびまたはルーティング接続処理ユニット 816 であり得る。

20

#### 【0100】

[0091]なおさらに別の構成では、ニューロンモデルは、共通特徴モデルをユーザに分散することと、共通特徴モデル上の分類子をトレーニングすることと、および／または分類子のうちの第1の分類子を第1のユーザに、第2の分類子を第2のユーザに分散することとのために構成される。ニューロンモデルは、共通特徴モデルを分散するための手段と、トレーニング手段と、複数の分類子のうちの第1の分類子を第1のユーザに、および第2の分類子を複数のユーザのうちの第2のユーザに分散するための手段（「分類子を分散するための手段」）とを含む。一態様では、共通特徴モデルを分散するための手段、トレーニング手段、および／または分類子を分散するための手段は、記載される機能を実行するように構成された汎用プロセッサ 502、プログラムメモリ 506、メモリブロック 504、メモリ 602、相互接続ネットワーク 604、処理ユニット 606、処理ユニット 704、ローカル処理ユニット 802、および／またはルーティング接続処理ユニット 816 であり得る。

30

#### 【0101】

[0092]さらなる構成では、ニューロンモデルは、第1の分類子モデルを学習するために、共通特徴マップのセットを第1の指定されたユーザからラベル付けされた例の第1のコーパスに適用することと、第2の分類子モデルを学習するために、共通特徴マップのセットを第2の指定されたユーザからラベル付けされた例の第2のコーパスに適用することと、および／または分類子モデルを分散することとのために構成される。ニューロンモデルは、第1の分類子モデルを学習するために、共通特徴マップのセットを第1の指定されたユーザからラベル付けされた例の第1のコーパスに適用するための手段と、第2の分類子モデルを学習するために、共通特徴マップのセットを第2の指定されたユーザからラベル付けされた例の第2のコーパスに適用するための手段と、および分散手段とを含む。一態様では、第1の分類子モデルを学習するために、共通特徴マップのセットを第1の指定されたユーザからラベル付けされた例の第1のコーパスに適用するための手段、第2の分類子モデルを学習するために、共通特徴マップのセットを第2の指定されたユーザからラベ

40

50

ル付けされた例の第2のコーパスに適用するための手段、および／または分散手段は、記載される機能を実行するように構成された汎用プロセッサ502、プログラムメモリ506、メモリブロック504、メモリ602、相互接続ネットワーク604、処理ユニット606、処理ユニット704、ローカル処理ユニット802、およびまたはルーティング接続処理ユニット816であり得る。

#### 【0102】

[0093]別の構成では、上述の手段は、上記の手段によって記載された機能を実行するように構成された任意のモジュールまたは任意の装置であり得る。

#### 【0103】

[0094]本開示のいくつかの態様によれば、各ローカル処理ユニット802は、ニューラルネットワークの所望の1つまたは複数の機能的特徴に基づいてニューラルネットワークのパラメータを決定して、決定されたパラメータがさらに適応され、同調され、更新されるにつれて、所望の機能的特徴に向けて1つまたは複数の機能的特徴を開発するように構成され得る。10

#### 【0104】

[0095]図9は、本開示の態様による、モデルを学習するための例示的なデータフロー900を示すブロック図である。図9を参照すると、ブロック902で、ニューラルネットワークは、初期の重みW0を有するモデルを学習するようにトレーニングされ得る。いくつかの態様では、ニューラルネットワークは、トレーニング画像のセット上のオブジェクト認識のためのモデルを学習するようにトレーニングされ得る。ニューラルネットワークは、たとえば、ディープニューラルネットワーク(DNN)を備え得る。DNNは、複数の隠れレイヤを有するニューラルネットワークである。20

#### 【0105】

[0096]ブロック904で、初期のモデル重み(「モデル」とも呼ばれる)W0は、ユーザ(たとえば、スマートフォンまたは他のデバイスなどのモデルデバイス)または他のエンティティにプッシュアウトされてもよく、分散されてもよい。いくつかの態様では、モデルは広く分散され得る(たとえば、1億または10億のデバイスのオーダー)。

#### 【0106】

[0097]ブロック906および908で、各モバイルデバイスは、特定のタスクを実行するためにモデルW0を使用し得る。たとえば、いくつかの態様では、モデルW0はモバイルデバイス上のデータの分類を提供し得る。たとえば、モデルW0は、デバイスユーザのために写真内のオブジェクトを識別および／またはラベル付けし得る。いくつかの態様では、オブジェクトは、モデルW0を使用して自動的に識別またはラベル付けされ得る。さらに、各モバイルデバイスは、写真が撮影されるときに、または、場合によっては、写真がプレビューされるときにモデルパラメータ更新を学習し得、モバイルデバイスiはまた、モデルパラメータ更新W0, iを計算および蓄積し得る。いくつかの態様では、デバイスiは、推論のために分散モデルW0のパラメータ(たとえば、重み)のみを使用し得、その更新をローカルに適用することはできない。

#### 【0107】

[0098]パラメータ更新W0, iは、いくつかの方法で計算され得る。たとえば、パラメータ更新W0, iは、ユーザにラベルを促し、逆伝搬を使用すること、または期間にわたってモデル内の1つのレイヤを対象とし、そのレイヤのための重み勾配をたとえばオートエンコーダ目的関数に基づいて計算することによって、計算され得る。もちろん、他のタイプの目的関数も使用され得る。たとえば、いくつかの態様では、スパース(sparse)オートエンコーダ、収縮オートエンコーダ、デノイジング(denoising)オートエンコーダ目的関数なども使用され得る。そのような目的関数は、正則化のペナルティで再構築を最小限に抑えることができる。パラメータ更新はまた、教師なし(unsupervised)ウェイクスリープロセス、または他の更新技法を使用して計算され得る。40

#### 【0108】

[0099]モバイルデバイス(たとえば、スマートフォン)は、ブロック910で、中央サ50

ーバ／ハブを介する収集のために、それらのモデル重み更新  $W_0, i$  を送信し得る。いくつかの態様では、モデル重み更新は、毎日、毎週、または毎月など、定期的に中央サーバに送信され得る。もちろん、これは単なる例示であり、限定するものではない。たとえば、いくつかの態様では、モバイルデバイスは、サーバからの要求に応答して更新を返送し得る（たとえば、サーバは更新のためにポーリングし得る）。別の例では、モバイルデバイスは、サーバ要求に応答して、または定期的にスケジューリングされた更新と組み合わせて、更新を送信し得る。さらに別の例では、更新はトレーニング例の蓄積に基づいて返送され得る（たとえば、最後に提供された更新以降の写真の時間、または最後に提供された更新以降の写真の数をとる）。

【0109】

10

[00100] ブロック 912 で、中央サーバ／ハブは次に、モバイルデバイスから受信したモデル重み更新  $W_0, i$  に基づいて新しいモデル  $W_1$  を計算し得る。

【0110】

[00101] いくつかの態様では、ブロック 914 で、新しいモデルは承認プロセスを介して承認され得る。ブロック 916 で、新しいモデル  $W_1$  は、モバイルデバイスユーザにパッショナイトされてもよく、分散されてもよい。ブロック 918 および 920 で、各モバイルデバイスは、特定のタスクを実行するためにモデル  $W_1$  を使用し得る。その後、モデルをさらに更新するために、プロセスが繰り返され得る。

モデル更新の計算

【0111】

20

[00102] 更新されたモデルは様々な方法で計算され得る。たとえば、いくつかの態様では、更新されたモデルは以下のように計算され得る。

【0112】

【数15】

$$W_{k+1} = W_k + \eta \left( \sum_i^n \Delta W_{kj} \right) \quad (15)$$

【0113】

上式で、 $n$  はユーザ更新の数であり、 $\eta$  は学習パラメータである。

30

【0114】

[00103] いくつかの態様では、重み更新はあらかじめ正規化され得る。たとえば、重み更新は、重み更新を返送する前に学習された写真の数によってあらかじめ正規化（分割）され得る。これは、モデル重みの直線平均を提供し得る。

【0115】

[00104] いくつかの態様では、更新もまた重み付けされ得る。一例では、重み更新は  $p_i$  の関数として重み付けされ得、画像の数は  $W_0, i$  を計算するために使用される。このように、数百枚の写真を撮ったユーザからの重み更新は、たとえば 1 枚の写真を撮ったユーザからの重み更新よりも大きな影響を与えることがある。したがって、この修正では、更新されたモデルは以下のように計算され得る。

【0116】

40

【数16】

$$W_{k+1} = W_k + \eta \left( \sum_i p_i \Delta W_{ki} \right) / \left( \sum_i p_i \right) \quad (16)$$

【0117】

[00105] 重み付けされた更新の場合、ユーザによって重み付けし過ぎることから保護することが望ましい場合がある。すなわち、たとえば、面白くない写真（たとえば、白い壁の多数の写真）を撮るユーザ、個々のユーザからの過度の表現、およびモデルを意図的に

50

劣化させようとする攻撃者に対して保護することが望ましい場合がある。1つの手法は、重み更新を実行する前に、写真の数  $p_i$  を  $1 - p_i / p_{max}$  に上限を定める、または制限する、あるいは  $p_i = \min(p_i, p_{max})$  と等しくすることである。この場合、攻撃者からの重み更新を平均し保護するために、複数のユーザ（たとえば、すべてのユーザ、またはピアグループなどのそれらのセグメント）からの集計または多数の重み更新を使用し得る。さらに、大きな要素値を有する重み更新  $W_k$ 、 $i$  がフィルタリングまたは正規化され得る。

#### モデル承認

##### 【0118】

[00106] ユーザにプッシュアウトまたは分散されたモデルは、新しいモデル更新を学習することに加えて、アクティブな推論をサポートし得るため、更新されたモデルを承認するために便利であり得る。たとえば、いくつかの態様では、新たに学習された重みが推論性能を過度に低下させないことを保証するために、モデル性能が承認され得る。一方、推論性能が過度に低下されると、補正動作が開始され得る。

10

##### 【0119】

[00107] いくつかの態様では、更新されたモデル性能は、承認データセット上で測定され得る。一例では、更新されたモデル性能は、オブジェクト認識の精度またはFスコアを計算することによって測定され得る。この例では、更新されたモデルは、承認性能があらかじめ定められた量（たとえば、定義されたパーセンテージ、または固定された差）を超えて低下しない場合にのみ、分散またはプッシュアウトされ得る。性能が目標量を超えて低下する場合、補正措置が実施され得る。たとえば、いくつかの態様では、モデル更新は、ある期間にわたって（たとえば、このラウンドにわたって）無視されてもよく、通知がユーザに送信されてもよい（たとえば、それらのデルタ重みをリセットするために、および／あるいは現在のモデルまたは前のモデルを使用するために）。

20

##### 【0120】

[00108] いくつかの態様では、後述するように、外れ値検出器は、たとえば除去／無視するためにユーザの重みのサブセットを識別し得る。次いで、更新されたモデルは、残りの重みに基づいて再計算され得る。更新されたモデルはまた、再試験および承認プロセスを受ける場合がある。モデルが依然として目標メトリックに合致しない場合、追加またはより限定的な外れ値フィルタが使用され得る。

30

##### 【0121】

[00109] いくつかの態様では、勾配方向の直線探索が使用され得る。たとえば、これは、異なる学習率を有するいくつかの潜在的な更新されたモデルを計算し、最良の承認性能を有するモデル、目標承認性能しきい値を満たす最大学習率を有するモデル、または承認性能結果の関数として選択されたモデルを使用することによって行われ得る。

40

##### 【0122】

[00110] 新しい、または更新されたモデルはまた、推論のために前のモデルを、および重み更新を計算するために新しいモデルを使用するための表示を含み得る。これは、推論性能に影響を与えるずにモデル性能を低下させる方向で、いくつかのステップを探索するために学習することを可能にし得る。

40

##### 【0123】

[00111] 他の態様では、ユーザデバイスは、2つのモデル（たとえば、 $W_0$  および  $W_1$ ）を維持し得る。たとえば、ユーザデバイスは、1つのモデル（たとえば、 $W_0$ ）と、モデル更新に対応するサーバからのデルタ（たとえば、 $W_1$ ）とを維持し得る。2つのモデルはユークリッド距離に近い場合があるため、デバイスは、2つの別々のモデルを記憶する際に使用されるよりも少ないメモリで2つのモデルを維持し得る。この手法では、いくつかのステップの後に、モデル性能が新しくより良い性能点に改善された場合、モデルはこの新しいモデルを使用する表示とともにプッシュアウトされる。モデル性能が改善しない場合、1つまたは複数の補正動作（たとえば、上述の補正動作）が採用され得る。

50

##### 【0124】

[00112]さらに、承認データセットが古くなっていないことを保証するために、承認データセットのサニティチェックが実行され得る（たとえば、特徴学習を推進する新しい電話機、自動車などの新しいオブジェクトが見つからず、適切に承認データセットを更新する）。

#### 外れ値検出

##### 【0125】

[00113]いくつかの構成では、問題または潜在的な攻撃を示す反復重み更新を伴う個々の重み更新またはユーザ／デバイスを検出するために、外れ値検出器フィルタが任意で含まれ得る。外れ値フィルタは、重み更新の母集団に対する重み更新を試験し得る。たとえば、所与の重み値に対する更新の分散が計算され得る。重み更新が標的偏差の目標数を超えている場合、それは外れ値として識別され得る。10

##### 【0126】

[00114]同様に、上記で参照した直線探索が使用される場合、勾配ベクトル方向は極小値を指すべきである。ユーザ勾配と母集団平均勾配との内積がしきい値未満である場合、それは外れ値としてマークされ得る。勾配または勾配の要素の大きさが母集団の標準偏差の数を超える場合、それもまた外れ値としてマークされ得る。他の統計的試験が使用され得る。さらに、母集団および分散は、含まれる外れ値として、被験勾配の有無にかかわらず計算され得る。

##### 【0127】

[00115]所与の勾配更新が外れ値として記録される場合、重み更新の現在のラウンドのために除外されてもよく、更新により小さい重みが与えられてもよい。ユーザが、外れ値重み更新を繰り返し提供するとしてフラグが立てられた場合、更新はさらなる調査のためにフラグが立てられてもよく、ユーザは意図的にまたは意図せずにモデル精度を攻撃する可能性のあるユーザとして永久にフラグが立てられてもよい。いくつかの態様では、フラグを立てられたユーザは、フラグを立てられていないユーザよりも少ない寄与で重み更新を追加し得る。たとえば、更新されたモデルは、20

##### 【0128】

##### 【数17】

$$W_{k+1} = W_k + \eta \left( \frac{1}{n_{\text{non-flagged}}} \sum_{i \in \text{non-flagged}} \Delta W_{ki} + \frac{\gamma}{n_{\text{flagged}}} \sum_{i \in \text{flagged}} \Delta W_{ki} \right) \quad (17) \quad \text{30}$$

##### 【0129】

として計算され得、上式で、 $\gamma < 1$  は、フラグを立てられたユーザにより小さい寄与を提供するためである。あるいは、重みは更新から除外され得る（たとえば、 $\gamma = 0$ ）。

##### 【0130】

[00116]逆に、外れ値検出器は、より新規な画像に基づいて、および／またはより新規な特徴を含む、更新を決定する際に助け得る。外れ値検出器はまた、そのような新規な画像を供給するユーザを識別する際に助け得る。さらに、新規な画像および／またはユーザが識別されると、外れ値識別子は、それらの画像、ユーザ、および／または特徴の重みを除き得る。40

#### アーキテクチャ更新

##### 【0131】

[00117]更新されたモデルはアーキテクチャ更新を含み得る。モデルは、学習が続く間に推論タスクのために利用され得るので、トレーニングデータが過剰適合しないようにするために、初期のトレーニングデータおよびデバイス特性に基づいて初期のモデル容量の大きさを決めるために有益な場合がある。しかしながら、分散された学習が進むにつれて、モデル容量によって性能が制限され得る。同時に、モデルが学習するにつれて、最下位レイヤはそれらが学習できるものに収束し始める（たとえば、第1のレイヤのガボール型エッジ検出器）。それらのレイヤにおける追加トレーニングは、潜在的可能性が限られて50

いる可能性がある。これに対処するために、モデルは成長および／または契約し得る。

#### 【0132】

[00118] 1つの手法では、現在のモデル容量の収束があるかどうかを決定するために、計算および承認段階中のモデル容量が監視され得る。収束メトリックは、エポックの目標数に性能が飽和したかどうかを決定するために、あるいは、重み更新の大きさ、しきい値よりも大きい要素の数の希薄さ、および／または勾配方向のコヒーレンスなどの重み更新の特徴を見るために、承認セットのモデル性能を追跡することを含む。トレーニングエポックおよび／またはトレーニング例の数はまた、モデルの成長および／または収縮が有益であるかどうかを決定するために使用され得る。

#### 【0133】

[00119] モデルの成長が示される場合、たとえば、1つまたは複数の既存のレイヤにおけるニューロンの数が増加されてもよく、1つまたは複数の追加のレイヤがモデルに追加されてもよい。新しいニューロンおよび／またはレベルは、いくつかの方法のうちの1つで追加され得る。

#### 【0134】

[00120] 1つの手法では、追加されたアーキテクチャ（たとえば、ニューロンおよび／またはレイヤ）は、即時の影響を有しないように追加され得る。たとえば、あるレイヤ内の新しいノードへの重みがゼロ（0）の値で構成されてもよく、および／または新しいレイヤが、識別関数を形成する重みのセットで構成されてもよい。次いで、後続の学習エポックが改良された重みを学習し始める。

#### 【0135】

[00121] 別の手法では、上述の手法で新しいレイヤが追加され得、次いで、重みのより良い初期セットを得るために、中央データトレーニングセットにトレーニング（教師あり、および／または教師なしトレーニング）が実施され得る。

#### 【0136】

[00122] 上記の手法はモデルの成長を可能にするが、モデルのプラスチック／学習可能部分を縮小するために同じ表示が使用され得る。たとえば、重み更新が上位レイヤに対してのみ計算および送信されるように、最下位レイヤは固定され得る。

分類子特徴学習者分割

#### 【0137】

[00123] 図10は、本開示の態様による、分類子1000のための例示的なアーキテクチャを示すブロック図である。分類子を構築するまでの1つの課題は、最初から学習を開始することなしに、クラスラベルを追加または減算する、またはより洗練された推論エンジンを構築する方法である。1つの観察は、ディープニューラルネットワークのようなモデルは、下位レイヤの学習特徴、および上位レイヤ内のそれらの特徴の推論エンジンとして見られ得るということである。さらに、特徴学習は多数のトレーニング例から最も恩恵を受ける可能性があるが、推論エンジンは、高品質特徴を使用している場合、多くのより少数の例で学習し得る。この観察を使用して、分散学習は特徴の重みなどのモデル更新を学習し、図10に示される中央データセットを使用して、最上位レイヤで毎回分類子を最初から再トレーニングすることができる。

#### 【0138】

[00124] 一例として、図10の分類子のための例示的なアーキテクチャ1000では、デバイスは入力データ1002を提供する（たとえば、写真を撮る、または他の感覚入力データを提供することができる）。モデル重み更新は、DNNの特徴レイヤとして分散学習特徴1004を提供するために、入力データ1002に基づいて計算され得る。次いで、デバイスは、集中的に学習された推論エンジン1006に（周期的に、または別の方法で）重み更新を送信し得、特徴モデルは、前述のように、これらの重み更新に基づいて更新され得る。その後、特徴モデルの重みが固定され得、集中的にラベル付けされたデータセット1008を使用して新しい画像分類子が特徴上でトレーニングされ得る。次いで、結果として得られるモデルは、オブジェクト認識能力の改善およびさらなる特徴モデル学

10

20

30

40

50

習のために、デバイスに分散またはプッシュアウトされ得る。

#### 【0139】

[00125]この拡張として、集中的に学習された推論エンジン1006は、オブジェクトラベルを追加、減算、結合すること、複数のオブジェクトへのラベル付けに拡張すること、または継続的に学習される特徴を利用して他の改善を提供することができる。オブジェクトラベルの追加、減算、または結合は、既存の画像の変更、および／または画像の追加／削除など、特徴上で推論エンジンを学習するために使用されるデータセットを適切に変更することによって行われ得る。

#### 【0140】

[00126]同様に、推論エンジン1006は分散学習された特徴1004の上に構築されるので、新しい推論プロセス、アーキテクチャ、または手法が使用され得る。たとえば、新しい推論プロセスは、画像ごとに単一のラベルではなく、画像ごとに複数のラベルを提供するようなラベル付け変更を含み得る。別の例では、人工ニューラルネットワーク(ANN)手法から、スパイキングニューラルネットワーク手法、サポートベクトルマシン(SVM)手法、または他の手法に切り替えることによって、推論エンジンアーキテクチャが修正され得る。

#### 【0141】

[00127]さらに、推論エンジンを集中的に学習することによって、異なるトレーニングデータセットを使用して特徴の同じセットに対して異なる推論エンジンをトレーニングすることによる、異なる使用ケース、デバイス、またはアプリケーションに対して異なる推論エンジンが学習され得る。

#### 【0142】

[00128]さらに別の拡張として、特徴重みを固定し、干渉エンジンのみをトレーニングする代わりに、特徴重みはまた、ユーザおよび／またはデバイスに分散され得る結果として得られるモデルを計算するために、1つまたは複数のトレーニングデータセット上で学習することからさらに改良され得る。逆に、推論エンジンは、特徴と同じ分散方法で学習され得る。

#### 【0143】

[00129]一構成では、ユーザが推論エンジンから1つまたは複数のモデルラベルを補正するときに、分散学習は、時々教師あり入力であるが、ほとんどが教師なしであり得る。この構成では、モデル重み更新を計算するために、複数の学習プロセスが使用され得る。また、この構成では、特徴／推論学習の概念分割による分散学習は、ユーザがモデル改善をより迅速に確認できるように、ユーザラベルを用いて推論エンジンをローカルに更新し得る。

#### 【0144】

[00130]さらに、プライバシー、オプトアウト、および／または帯域幅の目的のために、いくつかの態様では、ユーザ画像が中央サーバに提供されない場合がある。この場合、新しいモデルがユーザデバイスで受信されたときに、ローカルに記憶された画像およびラベルに基づいて重みを更新することによって推論エンジンが自動的に改良され得るように、画像はローカルラベルを有するユーザのデバイスにキャッシュされ得る。これは、モデルが分散された方法で学習し続けている間に、ユーザがラベル補正に基づいて更新を保持する推論エンジンを有することを可能にし得る。

#### 【0145】

[00131]図11は、本開示の態様による、モデルを学習するための例示的なデータフロー1100を示すロック図である。図11を参照すると、初期モデル $W_0$ は、ロック1102で、中央データセット上でトレーニングされ得る。ロック1104で、初期モデル $W_0$ は、たとえば、ユーザ1およびユーザ2にプッシュアウトまたは分散され得る。もちろん、これは単なる例示であり、モデルは任意の数のユーザ、ユーザまたは他のエンティティのグループに分散され得る。

#### 【0146】

10

20

30

40

50

[00132] 初期モデル $W_0$ は、各ユーザデバイス(1106、1108)でモデルを認識および学習するために使用され得る。しかしながら、学習された重み更新は、中央サーバに非同期的に提供され得る。すなわち、各ユーザは、重み更新を中央サーバに非同期的に送信し得る(1106および1108)。中央サーバが単一のモデル更新を受信すると(たとえば、ブロック1110でユーザ1から、またはブロック1116でユーザ2から)、または、場合によってはある期間にわたってユーザのサブセットからのモデル更新のセットを受信すると、サーバは新しいモデルを計算し、それをユーザに分散またはプッシュアウトし得る(1112、1118)。

## 【0147】

[00133] いくつかの態様では、新しいモデル更新は、更新を提供するユーザのサブセットにのみ送信され得る。すなわち、更新は異なって分散され得る。たとえば、ブロック1110で、中央サーバはユーザ1から重み更新を収集し、次に新しいモデル $W_1$ を計算する。ブロック1112で、中央サーバは $W_1$ をユーザ1にのみプッシュアウトまたは分散し得る。更新は以下のように表され得る。

## 【0148】

## 【数18】

$$W_1 = W_0 + \eta \left( \frac{1}{n} \sum_i \Delta W_{ki} \right) \quad (18)$$

## 【0149】

上式で、合計は、サブセット内の1人または複数のユーザに及ぶ。ブロック1114で、新しいモデル $W_1$ は、ユーザ1デバイスにおける認識および学習のために使用され得る。

## 【0150】

[00134] ブロック1118で、中央サーバは、別のユーザ(たとえば、ユーザ2から)、またはユーザのサブセットからモデル更新のセットを受信し、別の新しいモデル更新を計算し得る(1116)。このモデル更新は、更新が $W_0$ などのより古いモデル上で計算されていても、進行中のモデルに次のように適用され得る。

## 【0151】

## 【数19】

$$W_2 = W_1 + \eta \left( \frac{1}{n} \sum_i \Delta W_{ki} \right) \quad (19)$$

## 【0152】

ブロック1118で、新しいモデル $W_2$ は、他のユーザ(たとえば、ユーザ2)、またはユーザのグループに分散され得る。ブロック1120で、新しいモデル $W_2$ は、ユーザ2デバイスでの認識および学習のために使用され得る。

## 【0153】

[00135] いくつかの態様では、モデル更新(たとえば、 $W_1$ または $W_2$ )が受信されると、それらがどのモデルに基づいていたかに関して表示でマーク付けされるか、構成され得る。更新が初期モデル(たとえば、 $W_0$ )からしきい値数の更新の後に受信された場合、そのような更新は古くなったものとして破棄され得る。

## 【0154】

[00136] 非同期更新の場合、非同期更新時間はいくつかの要因に基づき得る。一例では、サーバとネットワークリソースとの負荷平衡を支援するために、ユーザの異なるサブセットに対して異なる時間に更新が計画され得る。第2の例では、更新は、計算されたローカルモデル更新の目標数(たとえば、写真の目標数)、またはデバイスから中央サーバに戻る高帯域幅チャネルの突然の利用可能性などの、いくつかのデバイスローカルメトリックが満たされた後に送信され得る。

10

20

30

40

50

## 【0155】

[00137] いくつかの態様では、学習された重み更新は、各学習計算（たとえば、写真）の後に更新を適用するオンライン方式か、目標数の学習計算（たとえば、写真）の後に更新を適用するミニバッチ処理を使用して、ローカルに適用され得る。たとえば、ユーザは、中央サーバから最後に受信されたモデル以降の総蓄積重み更新を返送し得る。この手法は、推論が非承認モデル上で実行されるために中間性能が低下するリスクを負って、ユーザがそれらのモデルをローカルに改善し、モデル空間のより多くをより迅速に探索することを可能にし得る。

## 【0156】

[00138] 2つのモデル（たとえば、推論値を報告するためのものと、モデル空間のより多くを学習するためのもの）をローカルに維持することによって、性能低下のリスクが低減され得る。もちろん、維持されるモデルの数は単なる例示であり、リソースの利用可能性に応じて任意の数のモデルが維持され得る。これは、モデルWおよび進行中の更新Wを維持し、Wを推論のために使用してW + Wをモデル学習のために使用することによって行われ得る。

10

## 【0157】

[00139] これらのモデル探索のケースでは、中央サーバは、前述の方法のようにモデル更新を適用し得る。いくつかの態様では、中央サーバは、どれがより良いモデル更新を提供するかを決定するために、承認セットに対して異なる更新を試験し得る。

20

## モデルパラメータの送信

## 【0158】

[00140] 分散モデル学習に関連付けられる1つの課題は、大容量モデルのモデルサイズが比較的大きくなる可能性があるため、モデルをプッシュアウトし、学習されたモデルパラメータを戻すという単純な手法は、多くの帯域幅を消費する可能性があるということである。さらに、中央サーバでは、多数のデバイス（たとえば、数億から数十億のデバイス）からモデル更新を受信すると、維持するために非常に大きなフローが生成される場合がある。デバイスにおける帯域幅およびメモリ利用を低減するために採用され得るいくつかの方法がある。

## デバイスから中央サーバへ

## 【0159】

30

[00141] 第1の手法は、各ユーザ（デバイス）が中央サーバに送信するWをサブサンプリングすることである。モデルが多数（たとえば、数百万または数十億）の重みパラメータを有する場合、Wベクトルは、その多くの要素を有する。数百万または数十億ものデバイスの各々が中央サーバに完全な重みベクトルを送信するのではなく、各ユーザ（デバイス）が、要素のサブセット（たとえば、要素のランダムなサブセット）を送信し得る。各W要素は、典型的には、誤差関数を最小化するように計算されるので、各要素の更新だけが良好な方向でなければならない。多数のデバイスが存在するため、適切な統計平均を使用するのではなく、すべてのユーザが彼らのすべての更新を送信すると、帯域幅は有效地に利用され得ない。一態様では、サーバは、モデルがユーザ（デバイス）にプッシュアウトされたときに返送するパラメータの数についてパラメータn pを送信し得る。ローカルユーザデバイスは、中央サーバに送信するために、Wベクトル内の対応する数の要素位置をランダムに選択し得る。このように、各学習更新において、ローカルデバイスは、目標とされるW要素を計算するために使用される中間値のみを計算し得る。さらに、ローカルユーザデバイスは、Wのn p要素を経時的に追跡するだけでよい。デバイスは、それらのn p要素を適切な時間に中央サーバに送信し得る。

40

## 【0160】

[00142] 次の反復の間に、他のオプションが実施され得る。たとえば、一構成では、デバイスは、要素位置の同じセットを維持してもよく、新しいランダム要素位置を再生成してもよい。さらに、中央サーバによってプッシュアウトされるパラメータn pの値は、たとえば、デバイスの数の増加、モデルサイズの変化、帯域幅の増加、および他の要因を考

50

慮するために、継時的に変化し得る。

【0161】

[00143]別の構成では、中央サーバは、デバイスからパラメータのすべてまたは一部を受信してもよく、モデルを更新するために使用される Wをサブサンプリングしてもよい。これは、更新ごとに中央サーバにおいて実行される計算量を制御するために行われ得る。いくつかの態様では、中央サーバはまた、受信したすべての更新からの Wのランダムなサブセットを使用し得る。他の態様では、中央サーバは、受信した更新の一部をドロップ(drop)し得る。

中央サーバからデバイスへ

【0162】

[00144]モデル更新をデバイスに分散またはプッシュアウトすることはまた、大量の帯域幅を消費する可能性がある。帯域幅を減少させるために実施され得るいくつかの手法がある。

【0163】

[00145]1つの手法は、モデルがすべてのユーザに共通であるため、ユーザに重み更新をブロードキャストまたはマルチキャストすることである。たとえば、一構成では、重み更新は、セルラーショートメッセージサービス(SMS)チャネル、ブロードキャストチャネル、またはオーバーヘッドロケーションなどのオーバーヘッドチャネルに分散され得る。

【0164】

[00146]別の手法は、前のモデル重み値に基づくアルゴリズムを使用してモデル重み更新を圧縮することである。たとえば、モデル  $W_{k+1}$  に対して、中央サーバは、 $W_{k+1} - W_k$  を計算し、次いで、モデル更新を送信するための小さい値を有するべきである、結果として得られるベクトルに対して標準圧縮プロセスを使用し得る。ユーザ(デバイス)が更新を受け取ると、デバイスはその更新を解凍して、それを前のモデルに追加し得る。代替で、 $p(W_{k+1} | W_k)$  の推定確率に基づいて、ハフマン圧縮が使用され得る。

【0165】

[00147]さらに、運動量を使用する DNNにおいて生じるような重み更新に相関がある場合、サーバは重み更新の二重の差を計算し得る。

【0166】

【数20】

$$(W_{k+1} - W_k) - (W_k - W_{k-1}) \quad (20)$$

【0167】

[00148]モデル更新のために、二重の差は圧縮されて送信され得る。いくつかの態様では、確率に基づく圧縮は  $p(W_{k+1} | W_k, W_{k-1})$  として使用され得る。

【0168】

[00149]さらに別の手法では、中央サーバは、ユーザデバイスが現在の反復に集中する(たとえば、更新する)ために、所与のモデルレイヤまたは重みのセットを示すことができる。この場合、サーバは、現在のモデル更新反復の対象とされる重みのセットを示すことができる。デバイスは、目標とされる重みのセットに関連する重み更新のみを追跡し得る。上記と同様に、デバイスは、この目標とされる重みのセットのランダムなサブセットをさらに選択し得る。デバイスは、反復の終了時に、それらのモデル重み更新を中央サーバに送信し得る。次に、サーバはこの重みのセットの更新されたモデルを計算し、次のモデル更新のためにこれらの更新された重みのみを送出し得る。いくつかの態様では、継時的な反復は、学習のために異なる重みセットを目標とし得る。

【0169】

[00150]上記の手法の拡張として、中央サーバは、いくつかの態様では、反復のために異なるレイヤまたは重みのサブセットを目標とするようにユーザの異なるサブセットに指示し得る。中央サーバはまた、どのレイヤまたはモデルサブセットが性能に最も大きな影

10

20

30

40

50

響を与えていたりするかをチェックするために承認チェック段階を使用し、それらの更新のみをプッシュアウトし得る。

#### 共通特徴にわたるカスタマイズされた分類子

##### 【0170】

[00151]本開示の態様は、共通特徴機能にわたるカスタマイズされた分類子をさらに対象とする。ユーザが野生のキノコをそれらの外観によって識別したく、ユーザがこの能力に対して支払うことを望むと仮定する。1人または複数のマッシュルームのエキスパートが、これらのエキスパートが自分の仕事に対して報酬を与えられ得るような方法で、彼らの知識をユーザおよび他のキノコ狩りをする人に伝える効率的な方法は何だろうか？キノコ以外にも、エキスパートラベリングの恩恵を受けることができるオブジェクトのクラスの他の例は、自動車、動物、美術品、医療診断画像などを含む。10

##### 【0171】

[00152]本開示の態様によれば、機械学習の力を利用する分類子が開示される。エンティティのセット（たとえば、指定されたユーザまたはエキスパート）は、ラベル付けされた例のコーパスを、中央サーバまたは「モデルストア」に提供し得る。この中央サーバはまた、特定の感覚様式（または、様式の組合せ）に関連する統計的特徴のセットを含み得る。これらの特徴は、教師なし方式で学習され得る。サーバは、分類子を計算するために、学習された特徴と、エキスパートが提供したラベル付けされた例のセットの両方を使用し得る。サーバは、ユーザが遭遇する様々なオブジェクトのクラスを彼らが計算することを可能にする、計算された分類子のパラメータをデバイスに分散し得る。20

##### 【0172】

[00153]分類子のパラメータを記憶するために消費されるメモリは、通常、完全なトレーニングコーパスよりも数桁小さい。このクライアント - サーバアーキテクチャはまた、2つのトレーニングコーパスの連結に関して単一の分類子をトレーニングし、異なるまたは重複するドメイン内の複数のエキスパートからの分類知識を統合する能力をユーザに与える可能性を許容する。

##### 【0173】

[00154]インフラストラクチャは、デバイスが適切な特徴のセット、分類子パラメータを記憶する能力、および正しい分類を実施するためにこれらのパラメータを展開する能力を有することを保証するために、サーバ側とユーザのデバイス側の両方で提供され得る。したがって、本開示のある態様では、潜在的に非常に多数のユーザの間で1人または複数の指定されたユーザまたはエキスパートの分類専門知識を共有するためのプロセスが開示される。1人または複数のユーザは、感覚データを収集するためにモバイルデバイスを使用し、場合によっては、これらのデータを有意なラベル（たとえば、キノコの種類を識別するためにカメラでキノコを見る）に分類することを望む場合がある。ラベル付けされたデータコーパスの形態であり得る「専門知識」は、1つまたは複数のサーバに供給され得、サーバは、それらの特徴にわたる分類子を計算するために、ラベル付けされたデータコーパスと前に学習された特徴のセットとを組み合わせ得る。次いで、サーバは、任意の関心のあるユーザのデバイスに分類子を分散し得る。サーバよりも多くのユーザが存在する場合がある。30

##### 【0174】

[00155]図12は、本開示の態様による、分類子を生成するための例示的なデータフローを示すフローチャートである。図12を参照すると、ブロック1102で、サーバは特徴Fのセットを1人または複数のユーザに分散し得る。

##### 【0175】

[00156]いくつかの態様では、ユーザは、入力データから同じ特徴を計算するための機構をそれぞれ有し得る。すべてのユーザが同じ入力特徴を共有することを保証するための1つの方法は、サーバが、バージョン番号とともにこれらの特徴をすべてのユーザにプッシュまたは分散することである。たとえば、1つまたは複数のレイヤを有する人工ニューラルネットワーク（ANN）は特徴を計算し得、したがって、このANNの接続重みをす40

べてのユーザに送信することによって、サーバはデータフローのこの部分を実行し得る。これらの特徴はまた、分散された教師なし学習を使用してまとめて学習され得る。

#### 【0176】

[00157]ユーザのセットに分類子を提供することに关心があるエンティティ、エキスパートまたは他の指定されたユーザは、ラベル付けされたデータをサーバに提供し得る。このラベル付けされたデータのコーパスは、入力特徴が変更された場合に分類子を再トレーニングできるように、サーバにとって利用可能なままであり得る。ブロック1204で、たとえば、エキスパートは、ラベル付けされた例のコーパス、データDおよびラベルLを含むEを、サーバに送信し得る。たとえば、ラベル付けされた例のコーパスは、画像のセット（および、各画像の一意のラベルのセットの形態であり得る。

10

#### 【0177】

[00158]ブロック1206で、サーバは、実施例Eのコーパスにおける各ラベル付けされたデータD<sub>i</sub>と、その対応するエキスパートが提供したラベルL<sub>i</sub>との間のマッピングを学習する分類子を構築または学習し得る。可能な分類子と学習方法には多くの選択肢がある。たとえば、一態様では、ANNは分類子に使用され得、トレーニングは逆伝搬を使用して行われ得る。

#### 【0178】

[00159]いくつかの態様では、複数のエキスパートのトレーニングセットに含まれる知識（たとえば、画像内のキノコと蝶の両方を識別し得る分類子）を組み合わせる分類子を構成することもできる。この組み合わせられた分類子は、たとえば、2つのトレーニングコーパスの混合を使用してトレーニングを実行することによって構築され得る。

20

#### 【0179】

[00160]追加の態様では、分類子パラメータは、アーキテクチャパラメータ（たとえば、ANNの場合、レイヤ内のユニットの数）を含み得る。これは、示唆または表示された所与のコーパスの複雑さが、より大容量を有する分類子を使用し、したがって、所与のレイヤ内により多くのユニットがある場合に有用であり得る。

#### 【0180】

[00161]一態様では、トレーニングは、固定された特徴F(d)上で分類子をトレーニングすることを含み得、または、たとえば、特徴レイヤおよび最上分類レイヤを通じて逆伝播することによって特徴を微調整することを含み得る。別の態様では、固定された特徴F(d)および/または更新されたF(d)からの重みデルタがユーザデバイスに送出され得る。別の態様では、トレーニングは、同じ共有された特徴F(d)上で2つのエキスパートトレーニングセットの2つの別個の分類子をトレーニングすることを含み得る。

30

#### 【0181】

[00162]さらに、分類子はまた、分類子が最上レベルまたは一般レベル、ならびにより特有の分類子(specific classifiers)を有するように、階層的に構成または組織され得る。たとえば、最上レベルの分類子は、画像を自動車として分類し得、より特有の分類子は、車の種類（たとえば、セダン、スポーツユーティリティ車、スポーツカーなど）を分類し得る。特有の分類子の複数のレイヤも提供され得る。たとえば、1つの特有なレイヤは、画像を1958年のフェラーリGTカリフォルニアスパイダーとして分類し得る。

40

#### 【0182】

[00163]いくつかの態様では、分類子はデータセットに適用され、最上レベルの分類を実行するように構成され得る。最上レベルの分類に基づいて、分類子は、サーバに1つまたは複数の追加の分類子を要求し得る。1つまたは複数の追加の分類子を受信すると、データセットに関するより特有な分類が実行され得る。

#### 【0183】

[00164]ブロック1208で、ユーザは、サーバ（たとえば、「モデルストア」）を介して選択肢の中央セットから知識または分類子の本体を選択し得る。ユーザは、どの分類子をダウンロードしたいかという選択をサーバに示し得る。これは、たとえば、ダウンロード可能なすべての分類子を表示するオンラインストアの形態を取り得る。このストアは

50

、分類子に加えて、分類子を計算するために低レベルの特徴をダウンロードするオプションをユーザに与え得る（たとえば、キノコ分類子の場合、ユーザは、自然画像についての低レベルの視覚的特徴のセットをまずダウンロードし得る）。ユーザはまた、ダウンロードされるべき複数の分類子を組み合わせられた分類子として指定し得る。

#### 【0184】

[00165]さらに、ユーザは、所望される分類子の階層内のレイヤを指定し得る。たとえば、ユーザは、リンゴなどの果物を分類するため的一般的な分類子が欲しい場合もあり、リンゴの種類（たとえば、グラニースミス、ピンクレディ、ふじ、ガラなど）をさらに区別し得る、より特有の分類子が欲しい場合もある。

#### 【0185】

[00166]ロック1210で、サーバは、要求された知識を、分類子Cを記述するパラメータの形態でユーザに提供し得る。ユーザが、構築および／またはダウンロードするための1つまたは複数の分類子を指定すると、サーバは、この分類子のパラメータをユーザのデバイスに分散またはプッシュし得る。ANNベースの分類子の場合、これらのパラメータは、たとえば、接続重みおよびバイアス項を備え得る。

#### 【0186】

[00167]いくつかの例では、サーバは、いくつかの分類子のパラメータまたは分類子の階層のレイヤを、自動的にユーザにプッシュし得る。これは、たとえば、ユーザを介して提供される感覚情報に基づき得る（たとえば、ユーザがスポーツカーの多数の画像を有する - ユーザがキャプチャスポーツカー画像をさらに分類することを可能にするために、より特有の分類子が提供され得る）。

#### 【0187】

[00168]ロック1212で、ユーザはデータdを収集し得る（たとえば、彼女のスマートフォンでキノコの写真を撮る）。dの特徴であるF(d)は、たとえば、前に提供された特徴Fのセットを使用してローカルに計算され得る。分類子Cは、未知の刺激（たとえば、キノコの種類）に対する推定されるエキスパートのラベルを計算するために、これらの特徴に適用され得る。

#### 【0188】

[00169]一旦、特徴Fのセットおよびダウンロードされた分類子Cを装備すると、ユーザは、分類c(F(d))を取得するために、データセットd（たとえば、キノコの画像）を収集し、その特徴F(d)を抽出し、これらを分類子に供給し得る。これらの特徴に関する分類子の出力は、エキスパートが前に提供したラベル付けされたコーパスEと一致するこの観察のクラスのエキスパートの意見を表し得る。本開示の態様によれば、多くの分類子（たとえば、比較的浅いANN）は比較的迅速に計算され得る。これは、データの取得直後に分類が行われ得、データ取得プロセスの一部としてすぐにユーザに提示され得ることを意味する。たとえば、ユーザのスマートフォンのカメラビューファインダは、推定されるキノコの種類を画像自体の上にリアルタイムで表示し得る。

#### 【0189】

[00170]あるいは、分類子が複雑である場合、ユーザのデータdの分類は、まずデバイス上で特徴F(d)を計算し、それらの特徴をサーバに送信することによって、再びサーバ上で実行され得る。次いで、サーバは、分類c(F(d))を計算し、その結果をユーザのデバイスに返送し得る。

ユーザフィードバック

#### 【0190】

[00171]ユーザが、彼らのデバイス上のデータを分類できる場合、システムに関連するフィードバックを提供したい場合がある。そのようなフィードバックは、たとえば、以下の形態を取り得る。

タイプ1：新しいラベル、ユーザが、分類子は所与の入力に対して誤ったラベルを生成したと信じ、正しいラベルが何であるべきかを知っている場合；

タイプ2：「間違ったラベル」メッセージ、ユーザが、分類子は所与の入力に対し

10

20

30

40

50

て誤ったラベルを生成したと信じるが、正しいラベルが何であるべきかを知らない場合；または、

タイプ3：分類子の初期結果に基づいて、ユーザが同じデータに対してより具体化された分類子を適用したい場合、異なる分類子をロードする要求。

#### 【0191】

[00172]フィードバックは、ユーザのデバイス、サーバ、エキスパートまたは指定されたユーザ、あるいはユーザまたは他のエンティティのグループに提供され得る。いくつかの様では、タイプ1のフィードバックは、プライベート分類子を構築するために使用され得る。たとえば、プライベート分類子は、ユーザが追加のラベル付けされた例を提供し得る、エキスパートが提供した分類子から導き出され得る。タイプ2のフィードバックは、ネガティブなラベル付けされた例を提供することによって分類子を再トレーニングするために、孤立して、または好ましくは他のユーザからのフィードバックと組み合わせて使用され得る。

#### 【0192】

[00173]タイプ3のフィードバックは、オブジェクトクラス間の関連付けのデータベースを他の分類子に構築するために使用され得る。たとえば、広範なオブジェクト分類の分類子を使用する人は、リンゴをイメージし、「apple」というラベルを受け取り、リンゴの特有の種類を決定するために、リンゴのためのより特有の分類子に切り替えることができる。この動作は、「apple」というラベルを提供する他の分類子が、同じ特有のリンゴ分類子に切り替えるオプションを自動的にユーザに提供することができるよう、フィードバックの形態で取り込まれ得る。そのようなタイプ3のフィードバックを蓄積することによって、システムは、複数の分類子を分類子の階層に順序付けまたは編成し、コンテキスト依存方法により特有の分類子への自動的な切替えを提供し得る。1つの分類子からより特有の分類子に切り替える決定は自動化され得、また、たとえば、ユーザがあるオブジェクトにとどまっている期間の長さ、または単一の画像内に存在するクラスのインスタンスの数に基づき得る（ある画像内に多数の「apple」がある場合、たとえば、リンゴのより特有の分類子が有用であり得る）。

モデルストア

#### 【0193】

[00174]これらの分類子およびエキスパートモデルのフロントエンドのクリアリング場所はモデルストアであり得る。モデルストアは、特定のユーザ（たとえば、エキスパート）が、彼らのラベル付けされたデータセットをアップロードし、それらのデータセットを使用して構築された分類子の価格を設定することを可能にし得る。モデルストアはまた、モデルをトレーニングし、共有される機能の効率を再利用するために、ユーザが上述のバックエンドプロセスを使用してモデルを購入することを可能にし得る。

#### 【0194】

[00175]モデルストア内の価格設定によって、各エキスパートがラベル付けしたデータセットに対する1回限りの価格設定が可能になってもよく、組合せ価格設定が可能になってもよい。組合せ価格設定の一例は、所与の特徴セット上の第1の分類子のためのより高い価格と、同じ特徴セットで構築された後続の分類子の低減された価格とを含み得る。代替的には、組合せ価格設定は、第1のキノコ分類子の所与の価格と、他のエキスパートからの後続のキノコ分類子の割引価格とを含み得る。

#### 【0195】

[00176]バックエンドは、増分価格を決定することを助けるために、追加のエキスパートがラベル付けしたデータを追加する段階的な改善のためのいくつかの共同性能スコアを計算し得る。モデルストアはまた、分類精度、ラベル付けされた画像の数など、どのエキスパートデータセットを購入するかをユーザが選択することを助けるためのメトリックを表示し得る。

#### 【0196】

[00177]モデルストアはまた、どのキノコ分類子が彼らのデータに最も適しているかを

10

20

30

40

50

評価するために、ユーザが、たとえば、彼らの電話機で取得したキノコのいくつかの画像から特徴 F ( d ) をアップロードすることを可能にし得る。ユーザからのサンプル画像上で最高の分類精度を達成するモデルが、購入するモデルになる。

#### 他の例示的な使用ケース

##### 【 0 1 9 7 】

[00178] いくつかの態様では、ユーザは、特徴の共通セットと組み合わせて動作する特有の分類子を「 a p p s t o r e 」または他のアプリケーション販売店から購入し得る。

##### 【 0 1 9 8 】

[00179] 一態様では、果物および野菜の粗い分類子は、ユーザのデバイスによって感知されているオブジェクトをリンゴとして識別し得る。さらに、オブジェクト上にとどまること、またはホバリングすることによって、分類子は、彼らがグラニースミスリンゴを見ていることをユーザに伝えるために、リンゴのより特有の分類子（たとえば、同じ共通特徴機能に対してトレーニングされる）をロードし得る。いくつかの態様では、分類子は、オブジェクトをさらに分類し得る 1 つまたは複数の他の分類子を識別し得る。

##### 【 0 1 9 9 】

[00180] 別の態様では、列車についてたくさん知っているユーザは、エキスパートの列車分類子を購入し、それを彼ら自身の知識で補うことができる。

##### 【 0 2 0 0 】

[00181] さらに別の態様では、D a v o s 社の着用する G o o g l e G l a s s に旅行しているユーザは、2つのエキスパート分類子（有名人の顔、およびチーズ）をヘッドアップディスプレイ用の 1 つの結合されたラベラに統合するカスタマイズされた分類子を作成し得る。

##### 【 0 2 0 1 】

[00182] なおさらに別の態様では、危険な場所においてカメラまたは他のセンサを装備したロボット群は、それらの環境（砂利、草、泥、瓦礫）のテクスチャを表現するための優れた視覚的特徴を発見するために、それらのカメラ（または他のセンサ）と教師なし学習とを使用し得る。さらに、加速度計とオドメータを使用して、いくつかのスカウトロボットは、ロボットがこの種の地形上で前進を行うことができるかどうかに基づいて、異なるテクスチャに「合格」と「不合格」のラベルを割り当てることができる。ロボットはまた、これらの特徴にわたるカスタム分類子を学習し得る。次いで、カスタム分類子は残りの群と共有され得る。

##### 【 0 2 0 2 】

[00183] 本開示の態様は、スパイキングニューロンおよびスパイキングニューロンモデルを記載したが、これは単なる例示であり、非スパイキングニューロンおよびニューロンモデルもまた使用され得る。さらに、本明細書に開示された概念および技法は、スパイキングと非スパイキング分散学習の両方に使用され得る。

##### 【 0 2 0 3 】

[00184] 図 1 3 は、本開示の態様による、モデルを学習するための方法 1 3 0 0 を示している。ブロック 1 3 0 2 で、プロセスは、1人または複数のユーザから 1 つまたは複数のモデル更新を受信する。ブロック 1 3 0 4 で、プロセスは、前のモデルおよびモデル更新に基づいて、更新されたモデルを計算する。さらに、ブロック 1 3 0 6 で、プロセスは、更新されたモデルに基づいて、更新されたモデルのサブセットに関連するデータを 1 人または複数のユーザに送信する。

##### 【 0 2 0 4 】

[00185] いくつかの態様では、更新されたモデルは、性能メトリックおよび / またはモデル容量に基づいて承認され得る。

##### 【 0 2 0 5 】

[00186] いくつかの態様では、更新されたモデルは、モデル更新の比較分析に基づいて外れ値を検出することに基づいて計算され得る。

10

20

30

40

50

## 【0206】

[00187]いくつかの態様では、更新されたモデルは、モデルアーキテクチャおよび／または学習率の変化を含み得る。アーキテクチャおよび／または学習率は、承認データに対するモデル性能および／または重み更新の希薄さに基づいて決定される。

## 【0207】

[00188]いくつかの態様では、サブセットは、モデルの新しくトレーニングされたレイヤのみを含み得る。いくつかの態様では、サブセットはモデルのランダムなサブセットを備え得る。

## 【0208】

[00189]図14は、本開示の態様による、モデルを学習するための方法1400を示している。ブロック1402で、プロセスは、共有された推論モデルに基づいてサーバからデータを受信する。ブロック1404で、プロセスは、受信データに基づいて1つまたは複数のモデルパラメータを含むモデルを生成する。ブロック1406で、プロセスは、モデルに基づいて推論を計算する。ブロック1408で、プロセスは、推論に基づいて1つまたは複数のモデルパラメータ更新を計算する。さらに、ブロック1410で、プロセスはモデルパラメータ更新に基づいてデータをサーバに送信する。

10

## 【0209】

[00190]いくつかの態様では、プロセスは、ローカルにキャッシュされたトレーニング例を使用して分類子をトレーニングすることをさらに含む。

20

## 【0210】

[00191]いくつかの態様では、データは、現在のモデル更新と前のモデル更新との間の差に基づいて送信され得る。たとえば、その差は、運動量モデルにおいて圧縮または使用され得る。

## 【0211】

[00192]いくつかの態様では、モデルパラメータ更新を計算すること、および／またはモデルパラメータ更新に基づいてデータを送信することは、計算および／または送信するためにモデルパラメータのランダムなサブセットを選択することを含む。

30

## 【0212】

[00193]図15は、本開示の態様による、分類子のセットを更新するための方法1500を示している。ブロック1502で、プロセスは、分類子の第1のセットをデータの第1のセットに適用する。データは、センサデータまたはユーザデバイスに記憶された他のデータを備え得る。さらに、ブロック1504で、プロセスは、遠隔デバイスに、分類子の第1のセットの出力、および／または分類子の第1のセットの適用の性能尺度に基づいて分類子更新を要求する。

## 【0213】

[00194]いくつかの態様では、要求はコンテキスト情報に基づき得る。コンテキスト情報は、たとえば、ユーザ入力情報、所定の期間（たとえば、1日、1週間、1カ月など）にわたる観察数、位置、活動、加速度計、バッテリ残量（たとえば、バッテリ寿命が短い場合、複雑度の低い分類子が示され得る）を含み得る。追加の態様では、要求は計算負荷に基づき得る。たとえば、計算負荷が高い（たとえば、所定のしきい値を上回る）場合、より複雑度の低い分類子が示され得る。一方、計算負荷が低い（たとえば、あらかじめ定められたしきい値よりも低い）場合、より複雑度の高い分類子が使用され得る。

40

## 【0214】

[00195]いくつかの態様では、性能尺度は、分類子の精度または信頼度、複数の分類子の一一致の表示、分類子の計算速度、および／または同等物を備え得る。

## 【0215】

[00196]図16は、本開示の態様による、分類子モデルを生成するための方法1600を示している。ブロック1602で、プロセスは、共通特徴モデルをユーザに分散する。ブロック1604で、プロセスは共通特徴モデル上の分類子をトレーニングする。さらに、ブロック1606で、プロセスは、第1の分類子を第1のユーザに、第2の分類子を第

50

2のユーザに分散する。

【0216】

[00197]いくつかの態様では、分類子のうちの1つまたは複数は、エンティティから取得されたラベル付けされたデータのセット上でトレーニングされ得る。エンティティは、ユーザ、特定の指定されたユーザまたは他のエンティティを備え得る。エンティティから取得されたラベル付けされたデータのセット上でトレーニングされた分類子ごとにメトリックが提供され得る。メトリックは、たとえば、分類精度に関する情報、またはラベル付けされた画像の数を含み得る。

【0217】

[00198]いくつかの態様では、プロセスは、遠隔デバイス上のデータから計算された1つまたは複数の特徴を受信する。さらに、プロセスは、1つまたは複数の特徴に基づいて遠隔デバイス上のデータを分類するための1つまたは分類子を決定する。次に、プロセスは、1つまたは複数の分類子の表示を遠隔デバイスに分散する。

【0218】

[00199]いくつかの態様では、プロセスは、遠隔デバイス上のデータから計算された特徴を受信する。プロセスはまた、受信した特徴に基づいて分類を計算する。さらに、プロセスは、分類を遠隔デバイスに送信する。

【0219】

[00200]いくつかの態様では、プロセスは、組み合わせられた分類子を生成するために、第1の分類子と第2の分類子とを組み合わせる。組合せ分類子は、ラベル付けされたデータの1つまたは複数のセットに関連付けられるクラスの分類を行うように構成され得る。プロセスはまた、組み合わせられた分類子をユーザのうちの1人または複数に分散する。

【0220】

[00201]図17は、本開示の態様による、分類子モデルを生成するための方法1700を示している。ブロック1702で、プロセスは、第1の分類子モデルを学習するために、共通特徴マップのセットを第1の指定されたユーザからラベル付けされた例の第1のコーパスに適用する。ブロック1704で、プロセスは、第2の分類子モデルを学習するために、共通特徴マップのセットを第2の指定されたユーザからラベル付けされた例の第2のコーパスに適用する。さらに、ブロック1706で、プロセスは、第1の分類子モデルと第2の分類子モデルとを含む分類子モデルを1人または複数のユーザに分散する。

【0221】

[00202]いくつかの態様では、組み合わせられた分類子が生成され得る。組み合わせられた分類子は、ラベル付けされた例のコーパス、ラベル付けされた例の追加のコーパスに基づいて、ならびに共通特徴機能の第1のセットおよび共通特徴機能の第2のセットを使用して生成され得る。このように、組み合わせられた分類子は、ラベル付けされた例の1つまたは複数のコーパス、および追加のコーパスに関連付けられるクラスの分類を行うように構成され得る。

【0222】

[00203]上述した方法の様々な動作は、対応する機能を実行することが可能な任意の好適な手段によって実行され得る。それらの手段は、限定はしないが、回路、特定用途向け集積回路(A S I C)、またはプロセッサを含む、様々なハードウェアおよび/またはソフトウェア構成要素および/またはモジュールを含み得る。概して、図に示されている動作がある場合、それらの動作は、同様の番号をもつ対応するカウンターパートのミーンズプラスファンクション構成要素を有し得る。

【0223】

[00204]本明細書で使用する「決定」という用語は、多種多様なアクションを包含する。たとえば、「決定」は、計算すること、算出すること、処理すること、導出すること、調査すること、ルックアップすること(たとえば、テーブル、データベースまたは別のデータ構造においてルックアップすること)、確認することなどを含み得る。さらに、「決

10

20

30

40

50

定」は、受信すること（たとえば、情報を受信すること）、アクセスすること（たとえば、メモリ中のデータにアクセスすること）などを含み得る。さらに、「決定」は、解決すること、選択すること、選定すること、確立することなどを含み得る。

#### 【0224】

[00205]本明細書で使用する、項目のリスト「のうちの少なくとも1つ」を指す句は、単一のメンバーを含む、それらの項目の任意の組合せを指す。一例として、「a、b、またはcのうちの少なくとも1つ」は、a、b、c、a-b、a-c、b-c、およびa-b-cを包含するものとする。

#### 【0225】

[00206]本開示に関連して説明した様々な例示的な論理プロック、モジュール、および回路は、汎用プロセッサ、デジタル信号プロセッサ（DSP）、特定用途向け集積回路（ASIC）、フィールドプログラマブルゲートアレイ信号（FPGA）または他のプログラマブル論理デバイス（PLD）、個別ゲートまたはトランジスタ論理、個別ハードウェア構成要素、あるいは本明細書で説明した機能を実行するように設計されたそれらの任意の組合せを用いて実装または実行され得る。汎用プロセッサはマイクロプロセッサであり得るが、代替として、プロセッサは、任意の市販のプロセッサ、コントローラ、マイクロコントローラまたは状態機械であり得る。プロセッサはまた、コンピューティングデバイスの組合せ、たとえば、DSPとマイクロプロセッサとの組合せ、複数のマイクロプロセッサ、DSPコアと連携する1つまたは複数のマイクロプロセッサ、あるいは任意の他のそのような構成として実装され得る。

10

20

#### 【0226】

[00207]本開示に関連して説明した方法またはアルゴリズムのステップは、ハードウェアで直接実施されるか、プロセッサによって実行されるソフトウェアモジュールで実施されるか、またはその2つの組合せで実施され得る。ソフトウェアモジュールは、当技術分野で知られている任意の形式の記憶媒体で存在し得る。使用され得る記憶媒体のいくつかの例は、ランダムアクセスメモリ（RAM）、読み出し専用メモリ（ROM）、フラッシュメモリ、消去可能プログラマブル読み出し専用メモリ（EEPROM）、電気的消去可能プログラマブル読み出し専用メモリ（EEPROM（登録商標））、レジスタ、ハードディスク、リムーバブルディスク、CD-ROMなどを含む。ソフトウェアモジュールは、単一の命令、または多数の命令を備えることができ、いくつかの異なるコードセグメント上で、異なるプログラム間で、複数の記憶媒体にわたって分散され得る。記憶媒体は、プロセッサがその記憶媒体から情報を読み取ることができ、その記憶媒体に情報を書き込むことができるよう、プロセッサに結合され得る。代替として、記憶媒体はプロセッサと一体化され得る。

30

#### 【0227】

[00208]本明細書で開示する方法は、説明した方法を達成するための1つまたは複数のステップまたはアクションを備える。本方法のステップおよび/またはアクションは、特許請求の範囲から逸脱することなく互いに交換され得る。言い換えれば、ステップまたはアクションの特定の順序が指定されない限り、特定のステップおよび/またはアクションの順序および/または使用は、特許請求の範囲から逸脱することなく変更され得る。

40

#### 【0228】

[00209]説明した機能は、ハードウェア、ソフトウェア、ファームウェア、またはそれらの任意の組合せで実装され得る。ハードウェアで実装される場合、例示的なハードウェア構成はデバイス中に処理システムを備え得る。処理システムは、バスアーキテクチャを用いて実装され得る。バスは、処理システムの特定の適用例および全体的な設計制約に応じて、任意の数の相互接続バスとブリッジとを含み得る。バスは、プロセッサと、機械可読媒体と、バスインターフェースとを含む様々な回路を互いにリンクし得る。バスインターフェースは、ネットワークアダプタを、特に、バスを介して処理システムに接続するために使用され得る。ネットワークアダプタは、信号処理機能を実装するために使用され得る。いくつかの態様では、ユーザインターフェース（たとえば、キーパッド、ディスプレ

50

イ、マウス、ジョイスティックなど)もバスに接続され得る。バスはまた、タイミングソース、周辺機器、電圧調整器、電力管理回路などの様々な他の回路にリンクし得るが、それらは当技術分野でよく知られており、したがってこれ以上は説明されない。

### 【0229】

[00210]プロセッサは、機械可読媒体に記憶されたソフトウェアの実行を含む、バスおよび一般的な処理を管理することを担当し得る。プロセッサは、1つまたは複数の汎用および/または専用プロセッサを用いて実装され得る。例としては、マイクロプロセッサ、マイクロコントローラ、DSPプロセッサ、およびソフトウェアを実行し得る他の回路を含む。ソフトウェアは、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語などの名称にかかわらず、命令、データ、またはそれらの任意の組合せを意味すると広く解釈されたい。機械可読媒体は、一例として、ランダムアクセスメモリ(RAM)、フラッシュメモリ、読み出し専用メモリ(ROM)、プログラマブル読み出し専用メモリ(PROM)、消去可能プログラマブル読み出し専用メモリ(E PROM)、電気的消去可能プログラム可能読み出し専用メモリ(EEPROM)、レジスタ、磁気ディスク、光ディスク、ハードドライブ、または他の任意の適切な記憶媒体、あるいはそれらの任意の組合せを含み得る。機械可読媒体はコンピュータプログラム製品において実施され得る。コンピュータプログラム製品はパッケージング材料を備え得る。

10

### 【0230】

[00211]ハードウェア実装形態では、機械可読媒体は、プロセッサとは別個の処理システムの一部であり得る。しかしながら、当業者なら容易に理解するように、機械可読媒体またはその任意の部分は処理システムの外部にあり得る。例として、機械可読媒体は、すべてバスインターフェースを介してプロセッサによってアクセスされ得る、伝送線路、データによって変調された搬送波、および/またはデバイスとは別個のコンピュータ製品を含み得る。代替的に、または追加で、機械可読媒体またはその任意の部分は、キャッシュおよび/または汎用レジスタファイルがそうであり得るように、プロセッサに統合され得る。論じた様々な構成要素は、ローカル構成要素などの特定の位置を有するものとして説明され得るが、それらはまた、分散コンピューティングシステムの一部として構成されているいくつかの構成要素などの様々な方法で構成され得る。

20

### 【0231】

[00212]処理システムは、すべて外部バスアーキテクチャを介して他のサポート回路と互いにリンクされる、プロセッサ機能を提供する1つまたは複数のマイクロプロセッサと、機械可読媒体の少なくとも一部分を提供する外部メモリとをもつ汎用処理システムとして構成され得る。あるいは、処理システムは、本明細書に記載のニューロンモデルとニューラルシステムのモデルとを実装するための1つまたは複数のニューロモーフィックプロセッサを備え得る。別の代替として、処理システムは、プロセッサを有する特定用途向け集積回路(ASIC)と、バスインターフェースと、ユーザインターフェースと、サポート回路と、単一のチップに統合された機械可読媒体の少なくとも一部とを用いて、あるいは1つまたは複数のフィールドプログラマブルゲートアレイ(FPGA)、プログラマブル論理デバイス(PLD)、コントローラ、状態機械、ゲート論理、個別ハードウェア構成要素、または他の任意の適切な回路、あるいは本開示全体を通じて説明した様々な機能を実行し得る回路の任意の組合せを用いて実装され得る。当業者なら、特定の適用例と、全体的なシステムに課される全体的な設計制約とに応じて、どのようにしたら処理システムについて説明した機能を最も良く実装し得るかを理解されよう。

30

40

### 【0232】

[00213]機械可読媒体はいくつかのソフトウェアモジュールを備え得る。ソフトウェアモジュールは、プロセッサによって実行されたときに、処理システムに様々な機能を実行させる命令を含む。ソフトウェアモジュールは、送信モジュールと受信モジュールとを含み得る。各ソフトウェアモジュールは、単一の記憶デバイス中に常駐するか、または複数の記憶デバイスにわたって分散され得る。例として、トリガイベントが発生したとき、ソフトウェアモジュールがハードドライブからRAMにロードされ得る。ソフトウェアモジ

50

ユールの実行中、プロセッサは、アクセス速度を高めるために、命令のいくつかをキャッシュにロードし得る。次いで、1つまたは複数のキャッシュラインが、プロセッサによる実行のために汎用レジスタファイルにロードされ得る。以下でソフトウェアモジュールの機能に言及する場合、そのような機能は、そのソフトウェアモジュールからの命令を実行したときにプロセッサによって実装されることが理解されよう。

#### 【0233】

[00214]ソフトウェアで実装される場合、機能は、1つまたは複数の命令またはコードとしてコンピュータ可読媒体上に記憶されるか、あるいはコンピュータ可読媒体を介して送信され得る。コンピュータ可読媒体は、ある場所から別の場所へのコンピュータプログラムの転送を可能にする任意の媒体を含む、コンピュータ記憶媒体と通信媒体の両方を含む。記憶媒体は、コンピュータによってアクセスされ得る任意の利用可能な媒体であり得る。限定ではなく例として、そのようなコンピュータ可読媒体は、RAM、ROM、EEPROM、CD-ROMまたは他の光ディスクストレージ、磁気ディスクストレージまたは他の磁気記憶デバイス、あるいは命令またはデータ構造の形態の所望のプログラムコードを搬送または記憶するために使用され得、コンピュータによってアクセスされ得る、任意の他の媒体を備えることができる。さらに、いかなる接続もコンピュータ可読媒体として適切に名づけられる。たとえば、ソフトウェアが、同軸ケーブル、光ファイバーケーブル、ツイストペア、デジタル加入者回線(DSL)、または赤外線(IR)、無線、およびマイクロ波などのワイヤレス技術を使用して、ウェブサイト、サーバ、または他のリモートソースから送信される場合、同軸ケーブル、光ファイバーケーブル、ツイストペア、DSL、または赤外線、無線、およびマイクロ波などのワイヤレス技術は、媒体の定義に含まれる。本明細書で使用するディスク(disk)およびディスク(disc)は、コンパクトディスク(disc)(CD)、レーザーディスク(登録商標)(disc)、光ディスク(disc)、デジタル多用途ディスク(disc)(DVD)、フロッピー(登録商標)ディスク(disk)、およびBlu-ray(登録商標)ディスク(disc)を含み、ディスク(disk)は、通常、データを磁気的に再生し、ディスク(disc)は、データをレーザーで光学的に再生する。したがって、いくつかの態様では、コンピュータ可読媒体は非一時的コンピュータ可読媒体(たとえば、有形媒体)を備え得る。さらに、他の態様では、コンピュータ可読媒体は一時的コンピュータ可読媒体(たとえば、信号)を備え得る。上記の組合せもコンピュータ可読媒体の範囲内に含まれるべきである。

#### 【0234】

[00215]したがって、いくつかの態様は、本明細書で提示する動作を実行するためのコンピュータプログラム製品を備え得る。たとえば、そのようなコンピュータプログラム製品は、本明細書で説明する動作を実行するために1つまたは複数のプロセッサによって実行可能である命令を記憶した(および/または符号化した)コンピュータ可読媒体を備え得る。いくつかの態様では、コンピュータプログラム製品はパッケージング材料を含み得る。

#### 【0235】

[00216]さらに、本明細書で説明した方法および技法を実行するためのモジュールおよび/または他の適切な手段は、適用可能な場合にユーザ端末および/または基地局によってダウンロードされ、および/または他の方法で取得され得ることを諒解されたい。たとえば、そのようなデバイスは、本明細書で説明した方法を実施するための手段の転送を可能にするためにサーバに結合され得る。代替的に、本明細書で説明した様々な方法は、ユーザ端末および/または基地局が記憶手段をデバイスに結合または提供すると様々な方法を得ることができるように、記憶手段(たとえば、RAM、ROM、コンパクトディスク(CD)またはフロッピーディスクなどの物理記憶媒体など)によって提供され得る。その上、本明細書で説明した方法および技法をデバイスに与えるための任意の他の好適な技法が利用され得る。

#### 【0236】

[00217]特許請求の範囲は、上記で示した厳密な構成および構成要素に限定されないこ

10

20

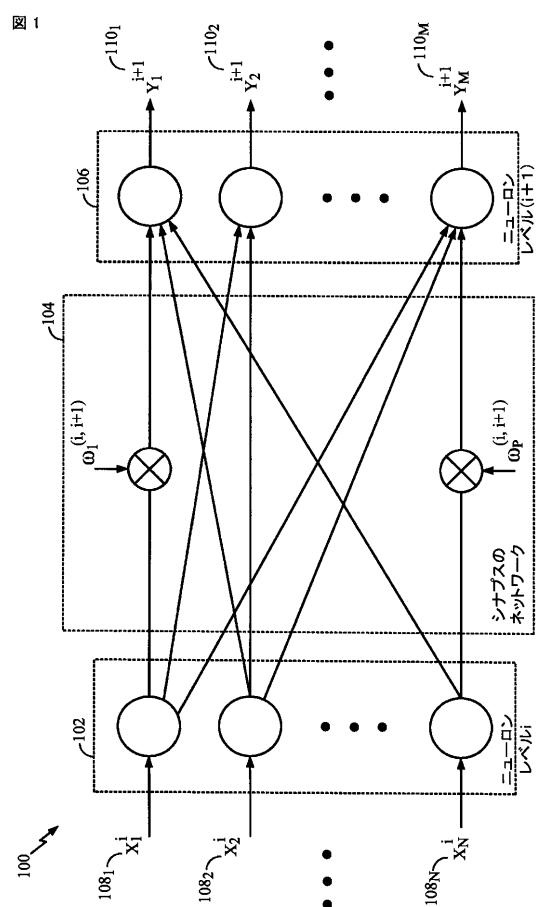
30

40

50

とを理解されたい。上記で説明した方法および装置の構成、動作および詳細において、特許請求の範囲から逸脱することなく、様々な改変、変更および変形が行われ得る。

【 四 1 】



【 四 2 】

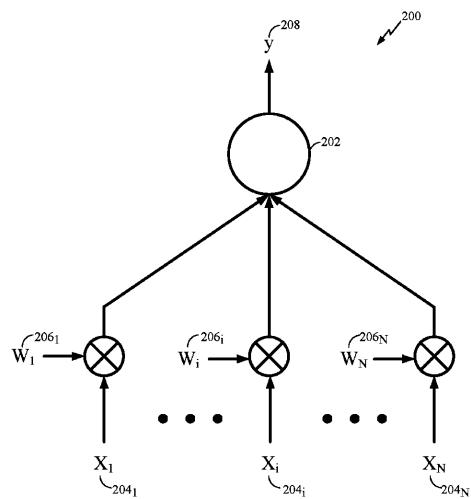


FIG. 1

【図3】

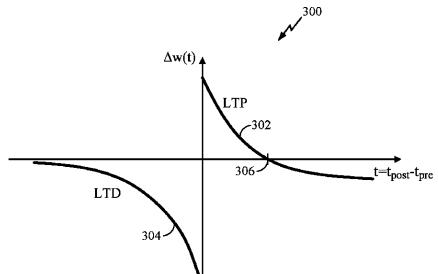


FIG. 3

【図4】

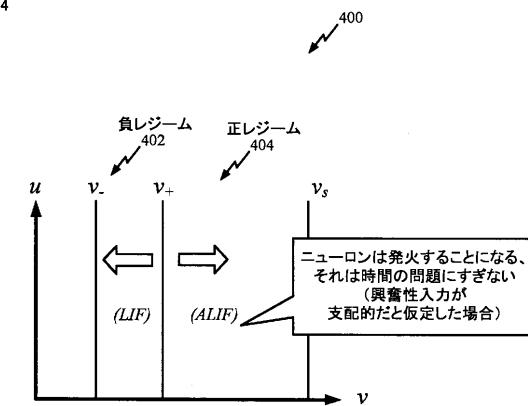


FIG. 4

【図5】

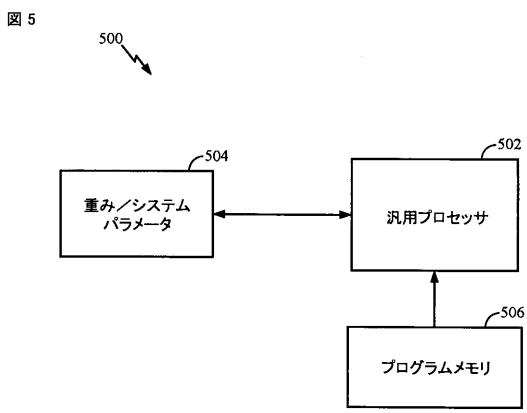


FIG. 5

【図6】

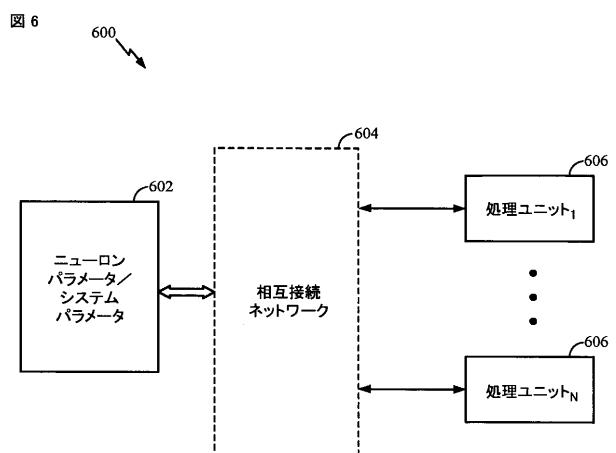


FIG. 6

【図7】

図7

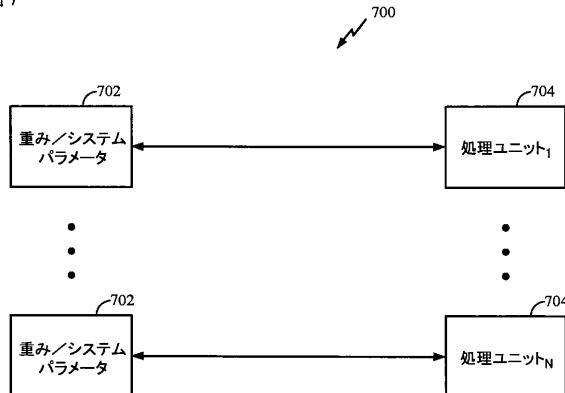


FIG. 7

【図8】

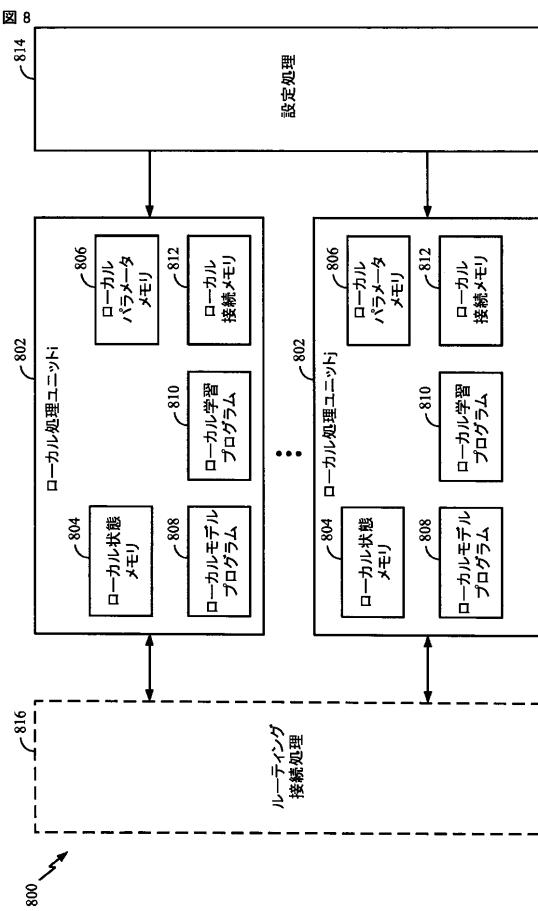


FIG. 8

【図9】

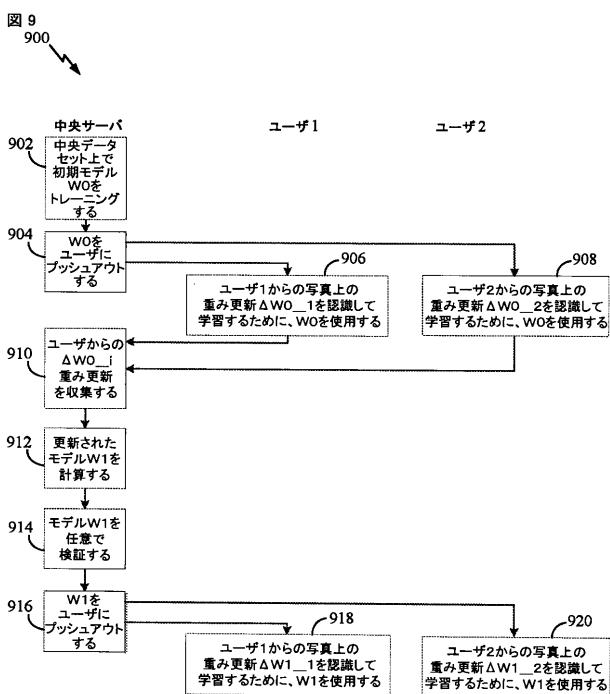


FIG. 9

【図10】

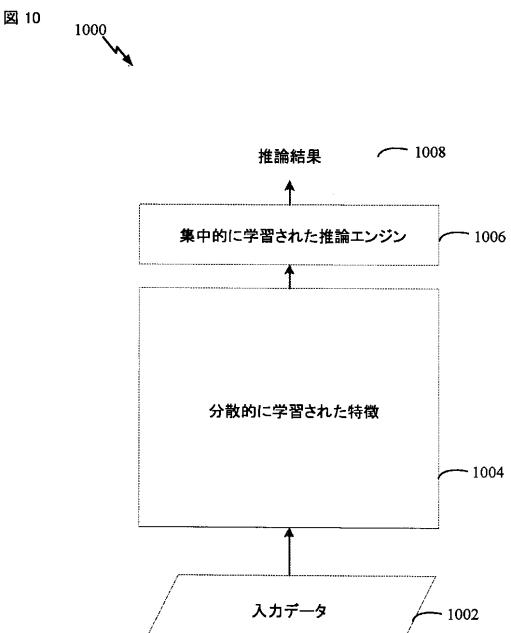


FIG. 10

【図 1 1】

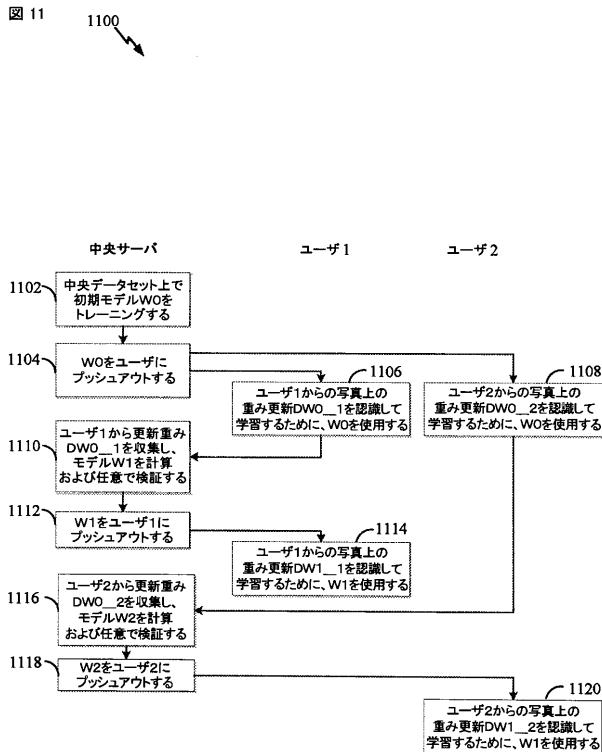


FIG. 11

【図 1 2】

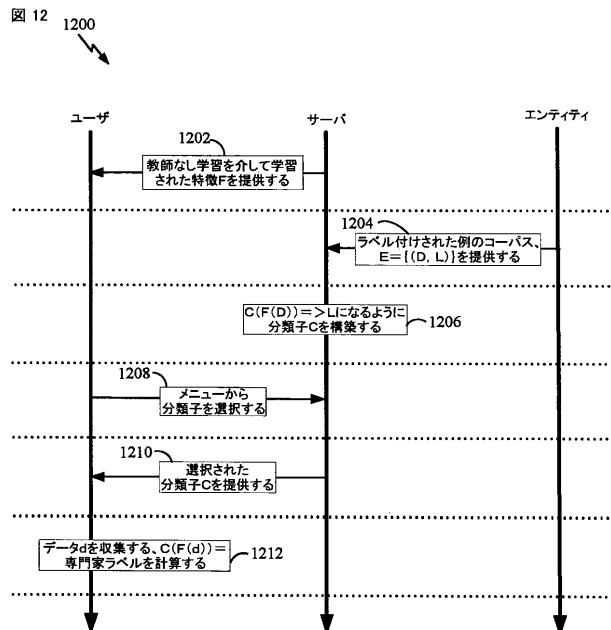


FIG. 12

【図 1 3】

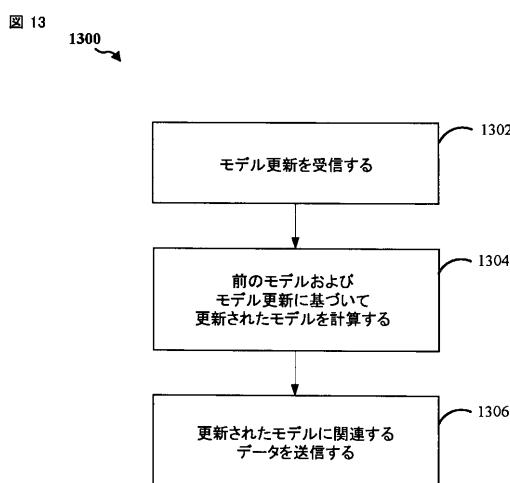


FIG. 13

【図 1 4】

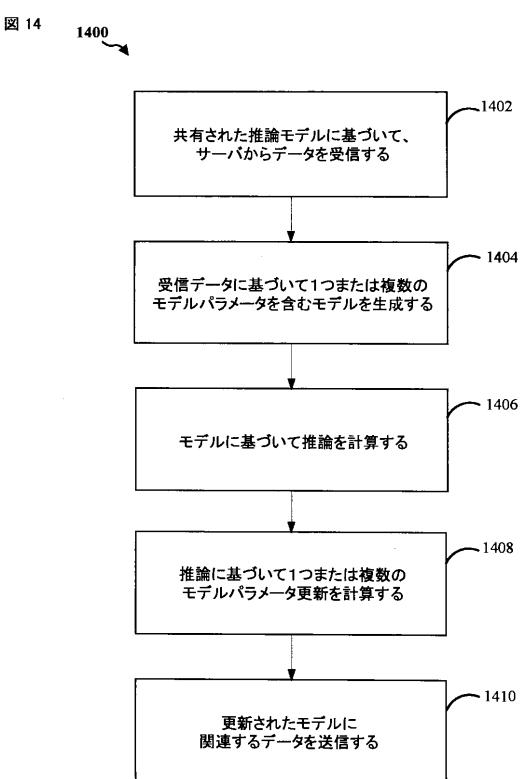
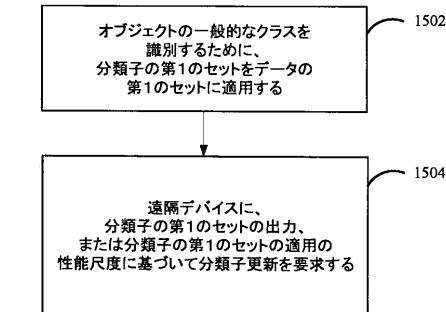


FIG. 14

【図 15】

図 15



【図 16】

図 16

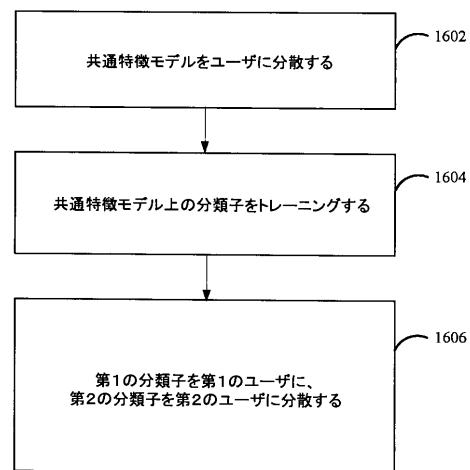


FIG. 15

FIG. 16

【図 17】

図 17

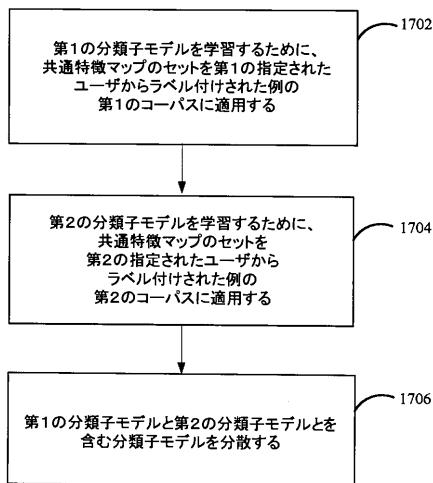


FIG. 17

**【手続補正書】**

【提出日】平成29年1月23日(2017.1.23)

**【手続補正1】**

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

**【補正の内容】**

【特許請求の範囲】

**【請求項1】**

分類子モデルを生成する方法であって、

共通特徴モデルを複数のユーザデバイスに分散することと、

前記共通特徴モデル上で複数の分類子をトレーニングすることと、

前記複数の分類子のうちの第1の分類子を前記複数のユーザデバイスのうちの第1のユーザに、および前記複数の分類子のうちの第2の分類子を前記複数のユーザデバイスのうちの第2のユーザに分散すること、前記第1の分類子は、前記第2の分類子とは異なる、と

を備える、方法。

**【請求項2】**

前記複数の分類子のうちの少なくとも1つは、少なくとも1つのエンティティから取得されたラベル付けされたデータのセット上でトレーニングされる、

請求項1に記載の方法。

**【請求項3】**

前記少なくとも1つのエンティティは、指定されたユーザを備える、

請求項2に記載の方法。

**【請求項4】**

前記少なくとも1つのエンティティから前記取得されたラベル付けされたデータのセット上で前記トレーニングされた分類子の各々にメトリックを提供することをさらに備える、

請求項2に記載の方法。

**【請求項5】**

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える、

請求項4に記載の方法。

**【請求項6】**

遠隔デバイス上のデータから計算された少なくとも1つの特徴を受信することと、

前記少なくとも1つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記データを分類するための少なくとも1つの分類子を決定することと、

少なくとも1つの分類子の表示を前記遠隔デバイスに分散することと

をさらに備える、請求項1に記載の方法。

**【請求項7】**

遠隔デバイス上のデータから計算された特徴を受信することと、

前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、

前記分類を前記遠隔デバイスに送信することと

をさらに備える、請求項1に記載の方法。

**【請求項8】**

ラベル付けされたデータの1つまたは複数のセットに関連付けられるクラスの分類を行うように構成された、組み合わされた分類子を生成するために、前記第1の分類子と前記第2の分類子を組み合わせることと、

前記組み合わされた分類子を前記複数のユーザデバイスのうちの少なくとも1人に分散することと

をさらに備える、請求項 1 に記載の方法。

【請求項 9】

前記複数の分類子のうちの少なくとも 1 つは、少なくとも 1 つのエンティティからのフィードバックに少なくとも部分的に基づいてトレーニングされる、

請求項 1 に記載の方法。

【請求項 10】

分類子モデルを生成するための装置であって、

メモリと、

前記メモリに結合された少なくとも 1 つのプロセッサと  
を備え、前記少なくとも 1 つのプロセッサは、

共通特徴モデルを複数のユーザデバイスに分散することと、

前記共通特徴モデル上で複数の分類子をトレーニングすることと、

前記複数の分類子のうちの第 1 の分類子を前記複数のユーザデバイスのうちの第 1 の  
ユーザに、および前記複数の分類子のうちの第 2 の分類子を前記複数のユーザデバイスの  
うちの第 2 のユーザに分散すること、前記第 1 の分類子は、前記第 2 の分類子とは異なる  
こと

を行なうように構成される、装置。

【請求項 11】

前記少なくとも 1 つのプロセッサは、前記複数の分類子のうちの少なくとも 1 つを、少  
なくとも 1 つのエンティティから取得されたラベル付けされたデータのセット上でトレ  
ニングすることを行うようにさらに構成される、

請求項 10 に記載の装置。

【請求項 12】

前記少なくとも 1 つのエンティティは、指定されたユーザを備える、

請求項 11 に記載の装置。

【請求項 13】

前記少なくとも 1 つのプロセッサは、前記少なくとも 1 つのエンティティから前記取得  
されたラベル付けされたデータのセット上で前記トレーニングされた分類子の各々にメト  
リックを提供することを行うようにさらに構成される、

請求項 11 に記載の装置。

【請求項 14】

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える  
、

請求項 13 に記載の装置。

【請求項 15】

前記少なくとも 1 つのプロセッサは、

遠隔デバイス上のデータから計算された少なくとも 1 つの特徴を受信することと、

前記少なくとも 1 つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記  
データを分類するための少なくとも 1 つの分類子を決定することと、

少なくとも 1 つの分類子の表示を前記遠隔デバイスに分散することと

を行うようにさらに構成される、請求項 10 に記載の装置。

【請求項 16】

前記少なくとも 1 つのプロセッサは、

遠隔デバイス上のデータから計算された特徴を受信することと、

前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、

前記分類を前記遠隔デバイスに送信することと

を行うようにさらに構成される、請求項 10 に記載の装置。

【請求項 17】

前記少なくとも 1 つのプロセッサは、

ラベル付けされたデータの 1 つまたは複数のセットに関連付けられるクラスの分類を行

うように構成された、組み合わされた分類子を生成するために、前記第1の分類子と前記第2の分類子を組み合わせることと、

前記組み合わされた分類子を前記複数のユーザデバイスのうちの少なくとも1人に分散することと

行うようにさらに構成される、請求項10に記載の装置。

【請求項18】

前記少なくとも1つのプロセッサは、少なくとも1つのエンティティからのフィードバックに少なくとも部分的に基づいて、前記複数の分類子のうちの少なくとも1つをトレーニングすることを行うようにさらに構成される、

請求項10に記載の装置。

【請求項19】

分類子モデルを生成するための装置であって、

共通特徴モデルを複数のユーザデバイスに分散するための手段と、

前記共通特徴モデル上で複数の分類子をトレーニングするための手段と、

前記複数の分類子のうちの第1の分類子を前記複数のユーザデバイスのうちの第1のユーザに、および前記複数の分類子のうちの第2の分類子を前記複数のユーザデバイスのうちの第2のユーザに分散するための手段、前記第1の分類子は、前記第2の分類子とは異なる、と

を備える、装置。

【請求項20】

分類子モデルを生成するためのプログラムコードを符号化した非一時的コンピュータ可読媒体であって、前記プログラムコードは、プロセッサによって実行され、

共通特徴モデルを複数のユーザデバイスに分散するためのプログラムコードと、

前記共通特徴モデル上で複数の分類子をトレーニングするためのプログラムコードと、

前記複数の分類子のうちの第1の分類子を前記複数のユーザデバイスのうちの第1のユーザに、および前記複数の分類子のうちの第2の分類子を前記複数のユーザデバイスのうちの第2のユーザに分散するためのプログラムコード、前記第1の分類子は、前記第2の分類子とは異なる、と

を備える、非一時的コンピュータ可読媒体。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0236

【補正方法】変更

【補正の内容】

【0236】

[00217]特許請求の範囲は、上記で示した厳密な構成および構成要素に限定されないことを理解されたい。上記で説明した方法および装置の構成、動作および詳細において、特許請求の範囲から逸脱することなく、様々な改変、変更および変形が行われ得る。

以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

[C1]

分類子モデルを生成する方法であって、

共通特徴モデルを複数のユーザに分散することと、

前記共通特徴モデル上で複数の分類子をトレーニングすることと、

前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに、および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザに分散することと

を備える、方法。

[C2]

前記複数の分類子のうちの少なくとも1つは、少なくとも1つのエンティティから取得されたラベル付けされたデータのセット上でトレーニングされる、

C 1 に記載の方法。

[ C 3 ]

前記少なくとも 1 つのエンティティは、前記複数のユーザのうちの指定されたユーザを備える、

C 2 に記載の方法。

[ C 4 ]

前記少なくとも 1 つのエンティティから前記取得されたラベル付けされたデータのセット上で前記トレーニングされた分類子の各々にメトリックを提供することをさらに備える、

C 2 に記載の方法。

[ C 5 ]

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える、

C 4 に記載の方法。

[ C 6 ]

遠隔デバイス上のデータから計算された少なくとも 1 つの特徴を受信することと、

前記少なくとも 1 つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記データを分類するための少なくとも 1 つの分類子を決定することと、

少なくとも 1 つの分類子の表示を前記遠隔デバイスに分散することと  
をさらに備える、C 1 に記載の方法。

[ C 7 ]

遠隔デバイス上のデータから計算された特徴を受信することと、

前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、

前記分類を前記遠隔デバイスに送信することと

をさらに備える、C 1 に記載の方法。

[ C 8 ]

ラベル付けされたデータの 1 つまたは複数のセットに関連付けられるクラスの分類を行うように構成された、組み合わされた分類子を生成するために、前記第 1 の分類子と前記第 2 の分類子を組み合わせることと、

前記組み合わされた分類子を前記複数のユーザのうちの少なくとも 1 人に分散することと

をさらに備える、C 1 に記載の方法。

[ C 9 ]

前記複数の分類子のうちの少なくとも 1 つは、少なくとも 1 つのエンティティからのフィードバックに少なくとも部分的に基づいてトレーニングされる、

C 1 に記載の方法。

[ C 10 ]

分類子モデルを生成するための装置であって、

メモリと、

前記メモリに結合された少なくとも 1 つのプロセッサと

を備え、前記少なくとも 1 つのプロセッサは、

共通特徴モデルを複数のユーザに分散することと、

前記共通特徴モデル上で複数の分類子をトレーニングすることと、

前記複数の分類子のうちの第 1 の分類子を前記複数のユーザのうちの第 1 のユーザに、および前記複数の分類子のうちの第 2 の分類子を前記複数のユーザのうちの第 2 のユーザに分散することと

を行うように構成される、装置。

[ C 11 ]

前記少なくとも 1 つのプロセッサは、前記複数の分類子のうちの少なくとも 1 つを、少なくとも 1 つのエンティティから取得されたラベル付けされたデータのセット上でトレ

ニングすることを行うようにさらに構成される、

C 10 に記載の装置。

[ C 12 ]

前記少なくとも 1 つのエンティティは、前記複数のユーザのうちの指定されたユーザを備える、

C 11 に記載の装置。

[ C 13 ]

前記少なくとも 1 つのプロセッサは、前記少なくとも 1 つのエンティティから前記取得されたラベル付けされたデータのセット上で前記トレーニングされた分類子の各々にメトリックを提供することを行うようにさらに構成される、

C 11 に記載の装置。

[ C 14 ]

前記メトリックは、分類精度、またはラベル付けされた画像の数に関する情報を備える、

C 13 に記載の装置。

[ C 15 ]

前記少なくとも 1 つのプロセッサは、遠隔デバイス上のデータから計算された少なくとも 1 つの特徴を受信することと、前記少なくとも 1 つの特徴に少なくとも部分的に基づいて、前記遠隔デバイス上の前記データを分類するための少なくとも 1 つの分類子を決定することと、少なくとも 1 つの分類子の表示を前記遠隔デバイスに分散することとを行うようにさらに構成される、C 10 に記載の装置。

[ C 16 ]

前記少なくとも 1 つのプロセッサは、遠隔デバイス上のデータから計算された特徴を受信することと、前記受信された特徴に少なくとも部分的に基づいて分類を計算することと、前記分類を前記遠隔デバイスに送信することとようさらに構成される、C 10 に記載の装置。

[ C 17 ]

前記少なくとも 1 つのプロセッサは、ラベル付けされたデータの 1 つまたは複数のセットに関連付けられるクラスの分類を行うように構成された、組み合わされた分類子を生成するために、前記第 1 の分類子と前記第 2 の分類子を組み合わせることと、前記組み合わされた分類子を前記複数のユーザのうちの少なくとも 1 人に分散することとを行うようにさらに構成される、C 10 に記載の装置。

[ C 18 ]

前記少なくとも 1 つのプロセッサは、少なくとも 1 つのエンティティからのフィードバックに少なくとも部分的に基づいて、前記複数の分類子のうちの少なくとも 1 つをトレーニングすることを行うようにさらに構成される、

C 10 に記載の装置。

[ C 19 ]

分類子モデルを生成するための装置であって、共通特徴モデルを複数のユーザに分散するための手段と、前記共通特徴モデル上で複数の分類子をトレーニングするための手段と、前記複数の分類子のうちの第 1 の分類子を前記複数のユーザのうちの第 1 のユーザに、および前記複数の分類子のうちの第 2 の分類子を前記複数のユーザのうちの第 2 のユーザに分散するための手段とを備える、装置。

[ C 20 ]

分類子モデルを生成するためのコンピュータプログラム製品であって、  
プログラムコードを符号化した非一時的コンピュータ可読媒体を備え、前記プログラム  
コードは、  
共通特徴モデルを複数のユーザに分散するためのプログラムコードと、  
前記共通特徴モデル上で複数の分類子をトレーニングするためのプログラムコードと、  
前記複数の分類子のうちの第1の分類子を前記複数のユーザのうちの第1のユーザに、  
および前記複数の分類子のうちの第2の分類子を前記複数のユーザのうちの第2のユーザ  
に分散するためのプログラムコードと  
を備える、コンピュータプログラム製品。

## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No PCT/US2015/026621
---

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G06N3/02 G06N99/00 ADD.										
According to International Patent Classification (IPC) or to both national classification and IPC										
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) G06N										
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched										
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal										
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">Category*</th> <th style="text-align: left; padding: 2px;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="text-align: left; padding: 2px;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">X</td> <td style="padding: 2px;">US 2012/259801 A1 (JI LEI [CN] ET AL) 11 October 2012 (2012-10-11) the whole document -----</td> <td style="padding: 2px;">1-20</td> </tr> <tr> <td style="padding: 2px;">X</td> <td style="padding: 2px;">LIXIN DUAN ET AL: "Domain Adaptation From Multiple Sources: A Domain-Dependent Regularization Approach", IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE, PISCATAWAY, NJ, USA, vol. 23, no. 3, 31 March 2012 (2012-03-31), , pages 504-518, XPO11424893, ISSN: 2162-237X, DOI: 10.1109/TNNLS.2011.2178556 the whole document ----- -/-</td> <td style="padding: 2px;">1-20</td> </tr> </tbody> </table>		Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	US 2012/259801 A1 (JI LEI [CN] ET AL) 11 October 2012 (2012-10-11) the whole document -----	1-20	X	LIXIN DUAN ET AL: "Domain Adaptation From Multiple Sources: A Domain-Dependent Regularization Approach", IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE, PISCATAWAY, NJ, USA, vol. 23, no. 3, 31 March 2012 (2012-03-31), , pages 504-518, XPO11424893, ISSN: 2162-237X, DOI: 10.1109/TNNLS.2011.2178556 the whole document ----- -/-	1-20
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.								
X	US 2012/259801 A1 (JI LEI [CN] ET AL) 11 October 2012 (2012-10-11) the whole document -----	1-20								
X	LIXIN DUAN ET AL: "Domain Adaptation From Multiple Sources: A Domain-Dependent Regularization Approach", IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE, PISCATAWAY, NJ, USA, vol. 23, no. 3, 31 March 2012 (2012-03-31), , pages 504-518, XPO11424893, ISSN: 2162-237X, DOI: 10.1109/TNNLS.2011.2178556 the whole document ----- -/-	1-20								
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.										
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed										
T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family										
Date of the actual completion of the international search	Date of mailing of the international search report									
17 July 2015	27/07/2015									
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentstaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Valencia, Erika									

**INTERNATIONAL SEARCH REPORT**

International application No PCT/US2015/026621
---

**C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>Jeffrey Dean ET AL: "Large scale distributed deep networks", The 26th annual conference on Neural Information Processing Systems (NIPS'25): 3-8 December 2012, 6 December 2012 (2012-12-06), XP055113684, Retrieved from the Internet: URL:<a href="http://books.nips.cc/papers/files/nips25/NIPS2012_0598.pdf">http://books.nips.cc/papers/files/nips25/NIPS2012_0598.pdf</a> [retrieved on 2014-04-11] the whole document</p> <p>-----</p>	1-20
A	<p>KR 2012 0053211 A (KOREA ELECTRONICS TELECOMM [KR]) 25 May 2012 (2012-05-25) abstract</p> <p>-----</p>	1-20

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No  
PCT/US2015/026621

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012259801 A1	11-10-2012	NONE	
KR 20120053211 A	25-05-2012	KR 20120053211 A US 2012124037 A1	25-05-2012 17-05-2012

Form PCT/ISA/210 (patent family annex) (April 2005)

---

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,R0,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,D0,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US

(72)発明者 ウィールジンスキー、カシミール・マシュー

アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775

(72)発明者 ジュリアン、デイビッド・ジョナサン

アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775

(72)発明者 サラ、アンソニー

アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775