

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号

特開2023-88289

(P2023-88289A)

(43)公開日 令和5年6月26日(2023.6.26)

(51)国際特許分類

G 0 6 N 20/00 (2019.01)

F I

G 0 6 N 20/00 1 3 0

審査請求 未請求 請求項の数 20 O L (全24頁)

(21)出願番号 特願2022-194939(P2022-194939)
 (22)出願日 令和4年12月6日(2022.12.6)
 (31)優先権主張番号 17/550,551
 (32)優先日 令和3年12月14日(2021.12.14)
 (33)優先権主張国・地域又は機関
 米国(US)

(特許庁注：以下のものは登録商標)
 1. SMALL TALK

(71)出願人 390009531
 インターナショナル・ビジネス・マシー
 ンズ・コーポレーション
 INTERNATIONAL BUSI
 NESS MACHINES CORPO
 RATION
 アメリカ合衆国10504 ニューヨー
 ク州 アーモンク ニュー オーチャード
 ロード
 New Orchard Road, A
 rmonk, New York 105
 04, United States of
 America
 (74)代理人 100112690
 弁理士 太佐 種一

最終頁に続く

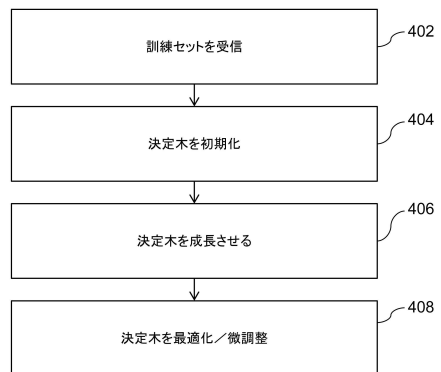
(54)【発明の名称】 機械学習において決定木を構築するコンピュータ実装方法、プログラム命令をその中に実装したコンピュータ可読記憶媒体を含むコンピュータプログラム製品、およびシステム(

(57)【要約】

【課題】次元削減による分類木と回帰木のパフォーマンス向上のためのコンピュータ実装方法、コンピュータプログラム製品、およびシステムを提供する。

【解決手段】機械学習のための決定木を構築および訓練するためのシステムおよび方法が提供され得る。訓練セットが受信され得る。決定木は、根ノードを構築することによって初期化されることができ、根ソルバは、訓練セットで訓練され得る。プロセッサは、決定木のノードを反復的に分割することによって決定木を成長させることができ、決定木のノードにおいて、ノードで受信された訓練セットのデータの特徴量に対して次元削減が実行され、次元削減されたデータが、決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割される。次元削減および分割は、非線形最適化問題を解くことに基づいて、ノードにおいて一緒に実行され得る。

【選択図】図4



【特許請求の範囲】**【請求項 1】**

機械学習において決定木を構築するコンピュータ実装方法あって、
訓練セットを受信することと、
根ノードを構築および前記訓練セットで根ソルバを訓練することによって、前記決定木を初期化することと、

前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させることと、

最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行することと、

を含む、コンピュータ実装方法。

【請求項 2】

前記決定木に対する所定のトポロジを受信することをさらに含み、

前記所定のトポロジが取得されるまで、前記ノードが反復的に分割される、請求項 1 に記載の方法。

【請求項 3】

前記決定木の前記葉ノードは、予測される目標値を返す前記ソルバを含む、請求項 1 に記載の方法。

【請求項 4】

前記決定木の前記葉ノードは、予測される目標値を返す回帰モデルを含む、請求項 1 に記載のコンピュータ実装方法。

【請求項 5】

正則化器を使用して前記決定木を最適化することをさらに含み、請求項 1 に記載のコンピュータ実装方法。

【請求項 6】

前記正則化器は、直交性正則化器を含む、請求項 5 に記載のコンピュータ実装方法。

【請求項 7】

前記正則化器は、多様化正則化器を含む、請求項 5 に記載のコンピュータ実装方法。

【請求項 8】

前記正則化器は、単一ルーティング正則化器を含む、請求項 5 に記載のコンピュータ実装方法。

【請求項 9】

前記決定木は、回帰問題を解決するために訓練される、請求項 1 に記載のコンピュータ実装方法。

【請求項 10】

前記決定木は、分類問題を解決するために訓練される、請求項 1 に記載のコンピュータ実装方法。

【請求項 11】

前記決定木の前記ノードは、少なくともルーティングノードおよび葉ノードを含み、前記次元削減は、前記ルーティングノードおよび葉ノードのそれぞれにおける最適化を伴って実行される、請求項 1 に記載のコンピュータ実装方法。

【請求項 12】

前記訓練セットは、不均衡データセットを含み、モデルの精度パフォーマンスの測定は、非線形メトリクスを含む、請求項 1 に記載のコンピュータ実装方法。

【請求項 13】

前記決定木は、前記訓練セットを受信することと、前記決定木を初期化することと、前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させることと、最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行することと、を含む、コンピュータ実装方法。

10

20

30

40

50

プログラム命令をその中に実装したコンピュータ可読記憶媒体を含むコンピュータプログラム製品であって、前記プログラム命令は、装置によって読み取り可能であり、

訓練セットを受信することと、

根ノードを構築および前記訓練セットで根ソルバを訓練することによって、決定木を初期化することと、

前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させることと、

最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行することと、

を前記装置に実行させる、

コンピュータプログラム製品。

【請求項 14】

所定のトポロジーが取得されるまで、前記ノードが反復的に分割される、請求項 13 に記載のコンピュータプログラム製品。

【請求項 15】

前記決定木の前記葉ノードは、予測される目標値を返す前記ソルバを含む、請求項 13 に記載のコンピュータプログラム製品。

【請求項 16】

前記決定木の前記葉ノードは、予測される目標値を返す回帰モデルを含む、請求項 13 に記載のコンピュータプログラム製品。

【請求項 17】

前記装置は、正則化器を使用して前記決定木を最適化する、請求項 13 に記載のコンピュータプログラム製品。

【請求項 18】

前記正則化器は、直交性正則化器、多様化正則化器、および単一ルーティング正則化器のうち少なくとも1つを含む、請求項 17 に記載のコンピュータプログラム製品。

【請求項 19】

プロセッサと、

前記プロセッサに結合されたメモリデバイスと、

を含み、

前記プロセッサは、少なくとも、

訓練セットを受信することと、

根ノードを構築および前記訓練セットで根ソルバを訓練することによって、決定木を初期化することと、

前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させることと、

最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行することと、

を行うように構成される、システム。

【請求項 20】

10
20
30
40
50

前記決定木の葉のノードは、予測される目標値を返す回帰モデルを含む、請求項 19 に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本願は、一般にコンピュータおよびコンピュータアプリケーションに関し、より詳細には、機械学習、次元削減技術を使用した分類木および回帰木の作成および訓練に関する。

【背景技術】

【0002】

決定木は、計算機的な魅力と様々なアプリケーションにおける強力なパフォーマンスで知られる機械学習モデルの人気クラスである。決定木は、データを再帰的に分割することによって生成されるデータの階層的なクラスタを学習することによって機能する。人気である一方、分類木および回帰木 (CART) のような最も基本的な方法は、分割を生成するために貪欲な基準やヒューリスティックに依存しており、計算コストを下げるために分割の最適性を犠牲にしている可能性がある。

10

【発明の概要】

【発明が解決しようとする課題】

【0003】

次元削減による分類木と回帰木のパフォーマンス向上のためのコンピュータ実装方法、コンピュータプログラム製品、およびシステムを提供する。

20

【課題を解決するための手段】

【0004】

本開示の概要は、例えば次元削減技術を使用して分類木および回帰木のパフォーマンスを高めるためのエンドツーエンドシステムのコンピュータシステムおよび方法の理解を助けるために与えられており、本開示または本発明を限定する意図で与えられているのではない。本開示の様々な態様および特徴は、有利には、いくつかの実施例において別々に、または他の実施例において本開示の他の態様および特徴と組み合わせて使用され得ることが理解されるべきである。したがって、異なる効果を達成するために、コンピュータシステムもしくはその動作方法またはその両方に変形および修正を加えることができる。

【0005】

30

機械学習における決定木を構築するコンピュータ実装方法は、一態様において、訓練セットを受信することを含むことができる。本方法はまた、根ノードを構築および前記訓練セットで根ソルバを訓練することによって、前記決定木を初期化することを含むことができる。本方法は、前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングするために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させることを含むことができる。本方法は、最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行することも含むことができる。

40

【0006】

一態様において、システムは、プロセッサと、前記プロセッサに結合されたメモリデバイスと、を含むことができる。プロセッサは、訓練セットを受信するように構成され得る。プロセッサは、根ノードを構築および前記訓練セットで根ソルバを訓練することによって、決定木を初期化するように構成され得る。プロセッサはまた、前記決定木のノードを反復的に分割することによって前記決定木を成長させることであって、前記決定木のノードにおいて、前記ノードで受信された前記訓練セットのデータの特徴量に対して次元削減が実行され、次元削減された前記データが、前記決定木の別のノードにルーティングする

50

ために、ルーティング関数に基づいて分割され、前記次元削減および前記分割が前記ノードにおいて一緒に実行され、前記決定木がルーティングノードおよび葉ノードを含む、成長させるように構成され得る。プロセッサはまた、最適化アルゴリズムによって、前記ルーティングノードにおけるルーティング関数、前記葉ノードにおけるソルバ、および前記決定木の各ノードでの次元削減の訓練を同時に実行するように構成され得る。

【0007】

本明細書に記載された1つまたは複数の方法を実行するために機械によって実行可能な命令のプログラムを記憶するコンピュータ可読記憶媒体も提供され得る。

【0008】

さらなる特徴、ならびに様々な実施形態の構造および動作を、添付の図面を参照して以下に詳細に説明する。図面において、同様の参照番号は、同一または機能的に類似の要素を示す。

10

【図面の簡単な説明】

【0009】

【図1】一実施形態における決定木学習を示す図である。

【図2】一実施形態における次元削減および決定木の訓練または学習を説明する別の図である。

【図3】一実施形態における一般的な決定木のフレームワークを説明する別の図である。

【図4】決定木の構築および学習の一実施形態における方法を示すフロー図である。

【図5】機械学習における決定木の構築および訓練を行うことができる一実施形態におけるシステムの構成要素を示す図である。

20

【図6】一実施形態によるシステムを実装し得る例示的なコンピュータまたは処理システムの概略図である。

【図7】一実施形態におけるクラウドコンピューティング環境を示す図である。

【図8】本開示の一実施形態におけるクラウドコンピューティング環境によって提供される一連の機能抽象化レイヤを示す図である。

【発明を実施するための形態】

【0010】

エンドツーエンドの木学習フレームワークが提供され得る。実施形態において、フレームワークは、非線形プログラミング技術を使用することができ、多数の特徴量およびサンプルを含み得る高次元マルチモーダル表形式データが存在し得るようなデータ設定において動作することができる。例えば、フレームワークは、非線形プログラム決定木フレームワークとすることができる。一態様において、フレームワークは、次元削減および訓練プロセスを一体化し、例えば、決定木の訓練または学習プロセスにおいて次元削減を統合する。一態様において、次元削減は、木ベースの教師あり学習モデルのサンプル外パフォーマンスを向上させる。また、このフレームワークは、そのパフォーマンスを向上させるために正則化器を実装してもよい。このフレームワークは、例えば、不均衡なデータセットと非線形な精度メトリクスを持つ分類タスクおよび回帰のタスクに適用することができる。また、このフレームワークは、分散訓練によるスケーラブルなアプローチを提供することができる。一実施形態において、フレームワークは、データの階層的クラスタリングのための方法を採用することができる。一実施形態において、分岐ルールは、確率モデルに基づくことができ、フレームワークは、下流学習モデルを認識することができる。モデルは、スケーラブルな分散削減確率勾配アルゴリズムによって訓練することができる。一実施形態において、フレームワークは、特徴量データのループ内の次元削減もしくは非線形正則化器を伴う主成分分析(PCA)に似た構造またはその両方を使用することができる。

30

40

【0011】

決定木は、回帰および分類において使用される学習モデルである。1つまたは複数の実施形態において、木構築に次元削減を組み込むエンドツーエンドの学習スキームを構築することができるシステムもしくは方法またはその両方を提示することができる。例えば、

50

決定木学習に次元削減を適用または統合することは、システムもしくは方法またはその両方が、最適な分類木フレームワークおよび回帰木フレームワークを計算的にスケーリングすることを可能にし得る。適切な次元削減を特定することで、決定木学習のパフォーマンスをさらに高めることができる。

【 0 0 1 2 】

本明細書に開示される決定木は、例えば、1つまたは複数のハードウェアプロセッサを含む、または1つまたは複数のハードウェアプロセッサと結合された1つまたは複数のコンピュータプロセッサ上でまたはそれによって実装、構築および訓練され得る。1つまたは複数のハードウェアプロセッサは、例えば、プログラマブルロジックデバイス、マイクロコントローラ、メモリデバイス、もしくは他のハードウェアコンポーネント、またはその両方などの構成要素を含んでもよく、これらは、本開示において説明されるそれぞれのタスクを実行するように構成されてもよい。結合されたメモリデバイスは、1つまたは複数のハードウェアプロセッサによって実行可能な命令を選択的に格納するように構成されてもよい。

10

【 0 0 1 3 】

プロセッサは、中央処理装置 (CPU)、グラフィック処理装置 (GPU)、フィールドプログラマブルゲートアレイ (FPGA)、特定用途向け集積回路 (ASIC)、他の適切な処理コンポーネントまたは装置、またはそれらの1つまたは複数の組み合わせであってもよい。プロセッサは、メモリデバイスと結合されてもよい。メモリデバイスは、ランダムアクセスメモリ (RAM)、読み取り専用メモリ (ROM) または別のメモリデバイスを含んでもよく、本明細書に記載される方法もしくはシステムまたはその両方に関連する様々な機能性を実装するためのデータもしくはプロセッサ命令またはその両方を格納してもよい。プロセッサは、メモリに格納された、または別のコンピュータ装置もしくは媒体から受信したコンピュータ命令を実行してもよい。

20

【 0 0 1 4 】

30

40

50

図1は、一実施形態における決定木を示す図である。一例として、あるデータを $\{(x_i, Y_i)\}_{i=1}^N$ として表現することができ、ここで回帰に対しては $Y_i \in \mathbb{R}$ 、分類に対しては $Y_i \in \{1, \dots, C\}$ であり、ここで Y_i はラベル（または予測されるラベル）を表し、 x_i は特徴量を表す。決定木は、根ノード（例えば、ノードA）、枝ノード（例えば、ノードB）、および葉ノード（例えば、ノード1、2、3）を含むことができる。データは、特徴量の値に応じて、枝ノードを介して異なる葉ノードに分割することができる。データポイントは、データがどこにフィッティングするかによって、左ノードまたは右ノードに送信されることができる。例えば、分割では、決定木のデータは、ノード n に対して、次のような形式の分割を経由してルーティングされる。データポイント x_i は、 $f_n(x_i) < b_n$ の場合、左にルーティングされ、データポイント x_i は、枝ルール関数 $f_n(x_i)$ に対して $f_n(x_i) \geq b_n$ の場合、右にルーティングされる。例えば、各分割は、データにフィッティングするようにデータから生成することができる。一実施形態では、決定木における枝ノードまたは根ノードにルーティングされたデータに対して、主成分分析（PCA）などの次元削減を実行することができる。ノード、例えば各分岐におけるモデルフィッティングで、減少した特徴量の数を見つけるために次元削減を適用することで、決定木の学習パフォーマンスを向上させることができる。例えば、次元削減は、ノードに送信されるデータのクラスタに対して実行することができる。一態様では、ノードに送信されるデータの特定のクラスタに対して次元削減を実行することにより、マルチモーダルなデータ型が存在し得るデータ設定に対応することができる。例えば、個々のクラスタに対して次元削減を実行することにより、データのマルチモーダル性を特定することができ、あるノードにおける1つのクラスタが、モデルフィッティングにおいて別のノードにおける別のクラスタとは異なる特徴量のセットを有し得る。

10

20

【0015】

30

図2は、一実施形態における次元削減および決定木の訓練または学習を説明する別の図である。一実施形態におけるエンドツーエンドの方法は、次元削減（例えば、PCA）と最適な決定木の学習を一体化されたモデルに組み合わせる。例えば、次元削減の最適化および予測モデルの学習の最適化（例えば、機械学習モデルなどにおける訓練損失の最小化）は、全てのノード（例えば、葉ノードおよび中間ノード）において同時に、または一度に、例えば、202に示す最適化を実行することにより、実施することができる。

$$\text{minimize } \sum_{k,i} L(f_k(x_{PCA_i}), y_i) \quad \text{s. t. 制約のセットが満たされる。}$$
ここで $f_k(x)$ は、 k 番目の葉ノードにおける予測モデルである。制約のセットは、解決される問題のタイプに応じて構成可能である。一実施形態において、木学習フレームワークのパフォーマンスを向上させるために、正則化器を導入することができる。

40

【0016】

一実施形態において、非線形プログラム決定木フレームワークは、多くの特徴量やサンプルが存在し得る高次元マルチモーダル表形式データに特化することができ、次元削減と訓練プロセスを一体化することができ、不均衡データセットと非線形精度メトリクスを有する分類タスクおよび回帰タスクに適用することができ、分散訓練でスケーラブルなアプ

50

ローチを形成することができる。

【 0 0 1 7 】

図 3 は、一実施形態における一般的な決定木のフレームワークを説明する別の図である。一態様において、決定木は、3つの構成要素を有し得る。ルータ：データを左または右にルーティングする r^i ノード、トランスフォーマー：データを低次元空間における新しい表現に変換する t^i ノード、ソルバ：それぞれの葉ノードにルーティングされた予測された所定のデータを返す s^i ノード。ルータ（例えば、302、304）は、あるノードから別のノードへデータ（例えば、306）を送信または分割する。トランスフォーマー（例えば、308、310、312、314、316）は、データに対して次元削減を適用することができる。ソルバ（例えば、320、322、324）は、ソルバに送信されるデータの異なるクラスにフィッティングされているモデルを含む。一実施形態では、各ソルバは、決定木の葉ノードにおいて異なるモデルとすることができる。

【 0 0 1 8 】

一実施形態において、トランスフォーマーは、PCAのような技術を採用することができる。例えば、PCAに基づく次元削減フレームワークは、学習における次元削減の影響を強調することができる。一実施形態において、正則化を使用して、分割およびルーティングを改善することができる。異なる正則化技術が採用され得る。例えば、一実施形態において、単一のルーティング正則化器は、各観測を単一のノードにルーティングすることを促進することができる。単一のルーティング正則化器は、葉ノードソルバにおける学習を改善することができ、ほぼ決定論的な決定木を出力することができる。単一のルーティング正則化器（例えば、 $-\gamma_2 \sum_{n=1}^N \sum_{\ell} \pi_{\ell}(\mathbf{x}^{(n)}; \Theta)^2$ ）は凹型のペナルティを持ち、 $\pi_{\ell}(\mathbf{x}^{(n)}; \Theta)$ を 0 または 1 に近づけることを促進する。

【 0 0 1 9 】

別の実施形態では、多様化正則化器（例えば、 $\gamma_1 \sum_{\ell} \left(\sum_{n=1}^N \pi_{\ell}(\mathbf{x}^{(n)}; \Theta) \right)^2$ ）は、データが葉ノード間で均等にルーティングされること、例えば、葉のそれぞれが等量の観測を受信することを促進することができる。多様化正則化器は、凸型ペナルティを有することができる。多様化正則化器は、観測データの階層的なクラスタリングを人為的に誘導することができ、モデルが、一般化しながらより最適に訓練データにフィッティングすることを可能にし、学習しながら同時にクラスタリングすることができる。別の実施形態では、直交性正則化器（ $\lambda \|C_n^T C_n - I_m\|_F$ ）は、次元削減（例えば、トランスフォーマー）をよりPCAに近づけるように促進することができる。例えば、直交性正則化器は、凸型ペナルティを有することができ、Cの列間の直交性を強制することにより、PCA構造を保持することが可能である。

【 0 0 2 0 】

一実施形態では、決定木学習フレームワークは、下流予測モデルを認識してルーティング関数にPCAのような投影を統合する。一般的な学習設定は、p個の特徴量を持つN個の観測データ $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \dots, (\mathbf{x}^{(N)}, \mathbf{y}^{(N)}) \in \mathcal{X} \times \mathcal{Y}$ （例えば、分類に対してはCクラス、または回帰に対しては $\mathbf{y} \in \mathbb{R}$ ）を含むことができる。

【 0 0 2 1 】

10

20

30

40

50

一実施形態では、フレームワークは、ルータが低次元空間における変換されたデータに基づいてデータを分割する、確率的決定木を統合する。例えば、決定木の学習スキームに投影構造を組み込むために、フレームワークは、確率的決定木または軟決定木を含んでもよい。一実施形態では、これらの木は、ある確率で x_i を左右にルーティングする。確率を決定する関数は、微分可能であってもよい。具体的には、例えば、ノード n は、確率

$$\sigma_n(f_n(x_i)) = \frac{1}{(1+e^{-f_n(x_i)})}$$

10

【 0 0 2 2 】

図3に示すように、一実施形態では、学習決定木の構成要素は、ルータノード 302、304 およびソルバノード 320、322、324 を含むことができる。一実施形態では、ルータノードは、各観測 $x^{(i)}$ を確率 p で左の子へルーティングし、確率 $1-p$ で右の子へルーティングする。一実施形態では、ルータは、ルーティングの前に、線形次元削減層を特徴量に適用してもよい。一実施形態では、木の葉ノードにおけるソルバノードが、予測されたターゲット $\hat{y}^{(i)}$ を返す。ソルバノードは、線形回帰、カーネル回帰、もしくは、その他、またはその組み合わせのような一般的な学習モデルであり得、形式的には、分類に対しては Δ_ℓ によってパラメータ化された $s_\ell(\mathbf{y}; \mathbf{x}, \Delta_\ell)$ 、回帰に対しては Δ_ℓ によってパラメータ化された $s_\ell(\mathbf{x}; \Delta_\ell)$ であり得る。

20

【 0 0 2 3 】

一実施形態では、トランスフォーマー 308、310、312、314、316 は、次元削減のためにノード間のアーク上に現れる。葉ノードではないノード（例えば、枝ノード）に現れるルータ 302、304 は、データを分割する。ソルバ 320、322、324 は、葉ノードに出現し、出力に対する推定値を生成する。フレームワークは、高次元のデータセットに対して機能することができる、次元削減に基づく特殊なモデルとすることができる。フレームワークは、訓練データとして N 個のラベル付きサンプル $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \dots, (\mathbf{x}^{(N)}, \mathbf{y}^{(N)}) \in \mathcal{X} \times \mathcal{Y}$ （回帰設定に対しては $\mathcal{X} \in \mathbb{R}^p$, $\mathcal{Y} \in \mathbb{R}$ 、分類設定に対しては $\mathcal{Y} \in \{1, \dots, C\}$ ）が与えられた場合に、分類設定に対しては $\mathbb{P}(\mathbf{y}|\mathbf{x})$ を、回帰設定に対しては $\mathbb{E}[\mathbf{y}|\mathbf{x}]$ を学習する。

30

【 0 0 2 4 】

40

50

一実施形態では、 n 番目のノードに対するルーティング関数は、 $r_n(\mathbf{x}; \Theta_n)$ と定義されることができ、それは、データポイント \mathbf{x} が左の子ノードにルーティングされる確率を示す。ルーティング関数は Θ_n によってパラメータ化され得る。一実施形態では、ルーティング関数は以下の通り選択され得る。

$$r_n(\mathbf{x}; \Theta_n) = \sigma(\mathbf{x}^\top \mathbf{C}_n \theta_n)$$

ここで、 $\mathbf{C}_n \in \mathbb{R}^{p \times m}$ 、 $\theta_n \in \mathbb{R}^{m \times 1}$ 、および $\sigma(z) = 1/(1 + e^{-z})$ は、シグモイド関数である。ここで、 p は例えば \mathbb{R}^p におけるようなサンプル \mathbf{x} に対する特徴量の数などの特徴量空間の次元を表し、 m は、新しい低次元特徴量空間に対する次元を表す。一実施形態において本明細書に記載される方法は、 \mathbb{R}^p における \mathbf{x} を例えば \mathbb{R}^m における低次元空間へ投影する。したがって、 m は調整可能なパラメータである。 \mathbf{C}_n を例えば \mathbb{R}^m におけるような低次元設定へ \mathbf{x} の投影としてみなすことができ、 θ_n を \mathbf{x} のより低い投影の分割パラメータとしてみなすことができる。例えば、行列 \mathbf{C}_n は次元削減に使用される。

10

【 0 0 2 5 】

一実施形態では、 ℓ 番目の葉ノードに対するソルバ関数は、 $\mathbb{P}(\mathbf{y}|\mathbf{x})$ に対する推定値を返す分類に対して $s_\ell(\mathbf{y}; \mathbf{x}, \Delta_\ell)$ として定義され得る。一実施形態では、回帰の場合、ソルバ関数は、 $\mathbb{E}[\mathbf{y}|\mathbf{x}]$ に対する推定値を返す $s_\ell(\mathbf{x}, \Delta_\ell)$ として定義され得る。ソルバ関数は、 Δ_ℓ によってパラメータ化され得る。分類の場合、フレームワークは、ソルバを多項ロジットモデルまたは定数に設定することができる。回帰の場合、ソルバは $\beta_\ell \in \mathbb{R}^p$ である $s_\ell(\mathbf{x}; \beta_\ell) = \mathbf{x}^\top \beta_\ell$ のような線形関数とすることができる。一実施形態において、ソルバの次元削減は、 $\mathbf{x}^\top \mathbf{C}_n$ の関数とすることができる。例えば、線形回帰ソルバでは $s_\ell(\mathbf{x}; \beta_\ell) = \mathbf{x}^\top \mathbf{C}_n \beta_\ell$ である。

20

【 0 0 2 6 】

推論を実行する、およびモデルを使用して \mathbf{x} に対する推定値 $\mathbb{P}(\mathbf{y}|\mathbf{x}; \Theta, \Delta)$ もしくは $\mathbb{E}[\mathbf{y}|\mathbf{x}; \Theta, \Delta)$ を生成するには、フレームワークは \mathcal{L}_ℓ および \mathcal{R}_ℓ を定義することができ、それらは \mathbf{x} を左右にルーティングし、それぞれ葉ノード ℓ に到達するノードのセットである。このことから、以下のことがわかる。

30

$$\mathbb{P}(\mathbf{y}|\mathbf{x}; \Theta, \Delta) = \sum_\ell (s_\ell(\mathbf{y}; \mathbf{x}, \Delta_\ell) [\prod_{n \in \mathcal{L}_\ell} r_n(\mathbf{x}; \Theta_n)] [\prod_{n \in \mathcal{R}_\ell} (1 - r_n(\mathbf{x}; \Theta_n))])$$

$$\mathbb{E}[\mathbf{y}|\mathbf{x}; \Theta, \Delta) = \sum_\ell (s_\ell(\mathbf{x}; \Delta_\ell) [\prod_{n \in \mathcal{L}_\ell} r_n(\mathbf{x}; \Theta_n)] [\prod_{n \in \mathcal{R}_\ell} (1 - r_n(\mathbf{x}; \Theta_n))]).$$

【 0 0 2 7 】

一実施形態において、分類問題では、フレームワークは、決定木を訓練するために最小化する目的関数として、負の対数尤度 (NLL) を利用することができる。

40

$$-\log \mathbb{P}(\mathbf{Y}|\mathbf{X}; \Theta, \Delta) = -\sum_{n=1}^N \log \mathbb{P}(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}; \Theta, \Delta)$$

【 0 0 2 8 】

50

一実施形態において、フレームワークは、目的関数に正則化器を追加することもできる。一例として、 C_n に対するPCAのような射影の場合、行列 C_n の列を直交させることができる。この制約を強制する方法として、正則化器 $\lambda\|C_n^T C_n - I_m\|_F$ を導入することができ、ここで I_m は $m \times m$ 単位行列である。一実施形態において、オーバーフィッティングを緩和し、多様性を促進するために、フレームワークは、1つのノードにルーティングされるデータポイントの数を減らすために正則化器を追加することができる。正則化器を定義するために、フレームワークは、以下を定義することができる。

$$\pi_\ell(\mathbf{x}; \boldsymbol{\Theta}) = \left[\prod_{n \in \mathcal{L}_\ell} r_n(\mathbf{x}; \boldsymbol{\Theta}_n) \right] \left[\prod_{n \in \mathcal{R}_\ell} (1 - r_n(\mathbf{x}; \boldsymbol{\Theta}_n)) \right] \quad 10$$

これはデータポイント \mathbf{x} が葉ノード ℓ にルーティングされる確率である。正則化項は次のようになる。

$$\gamma \sum_{n=1}^N \sum_\ell \pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta})^2.$$

これは、データポイントが1つの葉ノードにしかルーティングされない場合に大きくなるため、ペナルティ項である。

訓練のための完全な最適化問題は以下のようになる。

$$\min_{\boldsymbol{\Theta}, \Delta} -\log \mathbb{P}(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}, \Delta) + \lambda \|C_n^T C_n - I_m\|_F + \gamma \sum_{n=1}^N \sum_\ell \pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta})^2. \quad (1) \quad 20$$

この最適化問題を解くプロセスは決定木の訓練と呼ばれる。最適化問題(1)を解くと、ルータ $r_n(\mathbf{x}; \boldsymbol{\Theta}_n)$ に対するモデルパラメータ $\boldsymbol{\Theta}_n$ およびソルバ $s_\ell(\mathbf{y}; \mathbf{x}, \Delta_\ell)$ に対する Δ_ℓ が取得され得る。

【 0 0 2 9 】

回帰問題では、フレームワークは二乗損失を最小化することができる。

$$\sum_{n=1}^N [\mathbf{y}^{(n)} - \mathbb{E}(\mathbf{y}|\mathbf{x}^{(n)}; \boldsymbol{\Theta}, \Delta)]^2.$$

同じ正則化器を適用すると、訓練のための完全な最適化問題は以下のようになる。

$$\min_{\boldsymbol{\Theta}, \Delta} \sum_{n=1}^N [\mathbf{y}^{(n)} - \mathbb{E}(\mathbf{y}|\mathbf{x}^{(n)}; \boldsymbol{\Theta}, \Delta)]^2 + \lambda \|C_n^T C_n - I_m\|_F + \gamma \sum_{n=1}^N \sum_\ell |\pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta}) - \epsilon|. \quad (2) \quad 30$$

$$\min_{\boldsymbol{\Theta}, \Delta} \sum_{n=1}^N [\mathbf{y}^{(n)} - \mathbb{E}(\mathbf{y}|\mathbf{x}^{(n)}; \boldsymbol{\Theta}, \Delta)]^2 + \lambda \|C_n^T C_n - I_m\|_F + \gamma_1 \sum_{n=1}^N (\sum_\ell \pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta}))^2 - \gamma_2 \sum_{n=1}^N \sum_\ell \pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta})^2. \quad (3) \quad 40$$

$$\min_{\boldsymbol{\Theta}, \Delta} \sum_{n=1}^N [\mathbf{y}^{(n)} - \mathbb{E}(\mathbf{y}|\mathbf{x}^{(n)}; \boldsymbol{\Theta}, \Delta)]^2 + \lambda \|C_n^T C_n - I_m\|_F + \lambda \sum_{n=1}^N (\sum_\ell \pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta}))^2 - \gamma \sum_{n=1}^N \sum_\ell (\pi_\ell(\mathbf{x}^{(n)}; \boldsymbol{\Theta}) - 0.5)^2. \quad (4)$$

最適化問題(1)を解いた後、ルータ $r_n(\mathbf{x}; \boldsymbol{\Theta}_n)$ に対するモデルパラメータ $\boldsymbol{\Theta}_n$ およびソルバ $s_\ell(\mathbf{x}, \Delta_\ell)$ に対する Δ_ℓ が取得され得る。

【 0 0 3 0 】

決定木を構築する一実施形態は、新たな分割が予測誤差を改善しなくなるまで、幅優先探索方式において葉ノードを反復的に分割する貪欲法を使用することができる。別の実施形態では、決定木は、所定のトポロジーを使用して構築され得る。一実施形態において、木構築は、初期化段階、成長段階、および微調整段階を含むことができる。初期化段階において、フレームワークは、訓練データを訓練セットと検証セットとに分割することができる。フレームワークは、根ノード 0 を構築することと、ソルバ $s_0(\Delta_0)$ (例えば、根ソルバと呼ばれる) を訓練セットで訓練することと、によって、木を初期化することができる。一旦 $s_0(\Delta_0)$ が訓練されると、フレームワークは、微調整フェーズまで根ソルバを固定する。

10

【 0 0 3 1 】

成長段階において、所定のトポロジーを使用する実施形態では、フレームワークが木の所定のトポロジーを取得するまで、フレームワークは葉ノードを反復的に分割する。所定のトポロジーは、木構造を固定し、例えば、木の深さを固定する。例えば、各層に対して、フレームワークは、1つのソルバを元の根ノードソルバに割り当ててもよく、分割するために、フレームワークは、訓練データのランダムサブセットを使用して2つの新しいソルバを訓練し、分割を最適化するためにルータをフィッティングさせてもよく、フレームワークは、ルーティング関数を最適化してもよい。訓練プロセスは、木の他の部分のモデルパラメータを固定しながら、新しく追加されたルータおよびソルバに対して局所的に行われる。例えば、葉ノード ℓ の分割の場合、フレームワークは右の子のソルバに訓練データのランダムサブセットをフィッティングさせてよく、左の子のソルバを $s_0(\Delta_0)$ となるように設定してよい。そして、フレームワークは、 Θ および Δ における他のすべてのパラメータを固定して保持することにより、ルーティング関数 $r_\ell(\mathbf{x}; \Theta_\ell)$ を局所的に最適化することができる。例えば、ルーティング関数は $r_\ell(\mathbf{x}; \Theta_\ell) = 1/(1 + e^{-\mathbf{x}^T C_n \theta_n})$ であってよく、ここで、

20

$C_n \in \mathbb{R}^{p \times m}$ および $\theta_n \in \mathbb{R}^{m \times 1}$ である。一態様では、ルーティング関数は根ノードまたは中間ノードにおいて定義される。ソルバは、葉ノードにおける予測モデルである。これらは決定木を訓練するための一体化された最適化問題 (例えば、問題 (1) または (4)) において組み合わせられる。一態様では、ルータとソルバの両方で次元削減が起こり得る。

30

【 0 0 3 2 】

一実施形態において、微調整段階で、フレームワークは、確率的勾配降下法 (SGD) を使用して、(例えば、葉ノードにおける) すべてのソルバおよびルーティングパラメータを大域的にフィッティングさせてもよく、フレームワークは、基本分類器アルゴリズムを使用して、各葉ノードの分類器を再フィッティングさせてもよい。例えば、一旦成長段階が終了すると、フレームワークは、分類設定のための問題または式 (1) および回帰設定のための問題または式 (4) を解くことによって、木を大域的に最適化してもよい。フレームワークは、Adam アルゴリズムなどの確率的勾配降下型の方法を使用して、それぞれの問題を解いてもよい。

40

【 0 0 3 3 】

50

一実施形態では、例えば、微調整段階の終了時に、追加の改良を実行することができる。例えば、葉ノードは、訓練誤差を最小化するために、ルーティングされたデータを使用して再フィッティングすることができる。ルーティングされたデータを使用して再フィッティングするために、線形回帰の重み付けバージョンを最小化することができる。具体的には、例えば、各葉ノード l に対して、フレームワークは以下を最小化することができる。

$$\min_{\Delta_\ell} \sum_{n=1}^N \pi_\ell(\mathbf{x}^{(n)}; \Theta) [\mathbf{y}^{(n)} - s_\ell(\mathbf{x}^{(n)}, \Delta_\ell)]^2.$$

10

【 0 0 3 4 】

別の実施形態では、フレームワークは、1つまたは複数の前処理もしくは教師なし学習またはその両方を適用してもよく、これにより、学習方法に対するパフォーマンスを向上させることができる。例えば、低分散特徴量をフィルタリングすることと、特徴量を標準化することと、PCAを適用することとによって、良好なパフォーマンスを達成することができる。例えば、このような前処理を適用することで、訓練誤差を減少させることができる。

【 0 0 3 5 】

実施された実験は、本明細書に開示されたフレームワークの方法論が、特徴量の削減を含まない従来の決定木訓練技術よりも、訓練精度およびテスト精度を向上させることを示している。例えば、200の特徴量を有する4020個のサンプルの訓練データおよび4020個のサンプルのテストデータが実験に使用された。各試行の特徴量の数は特徴量削減因子によって削減された。特徴量の数が少ないほどパフォーマンスが向上することが確認された。

20

【 0 0 3 6 】

図4は、決定木の構築および学習の一実施形態における方法を示すフロー図である。本方法は、例えば、1つまたは複数のハードウェアプロセッサを含む、1つまたは複数のコンピュータプロセッサ上で実装または実行することができる。決定木は、各質問が特徴量および分割点を含む質問のシーケンスを学習する。402で、訓練セットが受信され得る。404で、根ノードを構築および訓練セットで根ソルバを訓練することによって、決定木が初期化され得る。406では、決定木のノードを反復的に分割することによって、決定木が成長され得る。例えば、決定木のノードにおいて、ノードで受信された訓練セットのデータの特徴量に対して次元削減が実行されることができ、削減された次元を有するデータが、決定木の別のノードにルーティングするために、ルーティング関数の最適化に基づいて分割されることができる。例えば、次元削減および分割は、ノードにおいて一緒に実行される。408において、決定木は、例えば、微調整されるなど、最適化されることができる。決定木は、ルーティングノードおよび葉ノードを含むことができ、本方法は、最適化アルゴリズムによって、ルーティングノードにおけるルーティング関数、葉ノードにおけるソルバ、および決定木の全てのノードでの次元削減の訓練を同時に (at the same time) または同時に (simultaneously) 実行することを含むことができる。

30

40

【 0 0 3 7 】

一実施形態において、決定木のための所定のトポロジーが受信されることができ、ここで、所定のトポロジーが取得されるまでノードを反復的に分割することができる。一実施形態において、決定木のノードは、少なくともルーティングノードおよび葉ノードを含むことができ、次元削減は、ルーティングノードおよび葉ノードのそれぞれにおける最適化を用いて実行され得る。一実施形態では、決定木の葉ノードは、予測された目標値を返すソルバを含むことができる。一実施形態において、決定木の葉ノードは、予測された目標値を返す回帰モデルを含むことができる。一実施形態において、決定木は、直交性正則化器、多様化正則化器、もしくは単一ルーティング正則化器またはその組み合わせなどの正則化器を使用して最適化され得る。一実施形態において、決定木は、回帰問題を解決する

50

ために訓練され得る。一実施形態において、決定木は、分類問題を解決するために訓練され得る。一実施形態において、訓練セットは、不均衡なデータセット（例えば、必ずしも均等に分割された目標値ではない）を含み得、モデルパフォーマンスメトリクスは、非線形の精度メトリクスを含み得る。例えば、非線形メトリクスは、F1スコア、マシューズ相関係数、およびFowlkes-Mallows indexを含み得る。

【0038】

一態様では、PCAと決定木を一体化されたフレームワークに統合することと、例えば、本明細書に開示された正則化器の1つまたは複数をエンドツーエンドの訓練に使用することとにより、予測精度を向上させることができる。確率的勾配降下のような問題（1）および（4）を解くための分散アルゴリズムを使用すると、訓練時間を削減することができる。一態様では、一体化されたフレームワークは、分類問題および回帰問題の両方で使用することができる。

10

【0039】

図5は、機械学習における決定木の構築および訓練を行うことができる一実施形態におけるシステムの構成要素を示す図である。中央処理装置（CPU）、グラフィックプロセッサユニット（GPU）、もしくはフィールドプログラマブルゲートアレイ（FPGA）、またはその組み合わせ、特定用途向け集積回路（ASIC）、もしくは別のプロセッサ、またはその両方などの1または複数のハードウェアプロセッサ502は、メモリデバイス504と結合されることができ、訓練セットに基づいて決定木モデルを生成および訓練することができる。未見データに基づく予測または質問への回答を作成することができる。メモリデバイス504は、ランダムアクセスメモリ（RAM）、読み取り専用メモリ（ROM）または別のメモリデバイスを含んでもよく、本明細書に記載される方法もしくはシステムまたはその両方に関連する様々な機能性を実装するためのデータもしくはプロセッサ命令またはその両方を格納してもよい。1または複数のプロセッサ502は、メモリ504に格納された、または別のコンピュータ装置もしくは媒体から受信したコンピュータ命令を実行してもよい。メモリデバイス504は、例えば、1または複数のハードウェアプロセッサ502を機能させるための命令もしくはデータまたはその両方を格納してもよく、オペレーティングシステムおよび命令もしくはデータまたはその両方の他のプログラムを含んでもよい。1または複数のハードウェアプロセッサ502は、訓練セットを含む入力を受信してもよい。例えば、少なくとも1つのハードウェアプロセッサ502は、機械学習における決定木モデルを生成してもよい。一態様において、かかる訓練データは、ストレージデバイス506に格納されてもよく、または遠隔装置からネットワークインタフェース508を介して受信されてもよく、決定木モデルを構築または生成するために、メモリデバイス504に一時的にロードされてもよい。学習された決定木モデルは、例えば、1または複数のハードウェアプロセッサ502によって実行するために、メモリデバイス504に格納されてもよい。1または複数のハードウェアプロセッサ502は、例えばネットワークを介してリモートシステムと通信するためのネットワークインタフェース508、およびキーボード、マウス、ディスプレイ、もしくはその他またはその組み合わせのような入力もしくは出力またはその両方のデバイスと通信するための入力/出力インタフェース510などのインタフェースデバイスと結合されてもよい。

20

30

40

【0040】

図6は、一実施形態におけるシステムを実装し得る例示的なコンピュータまたは処理システムの概略図である。コンピュータシステムは、適切な処理システムの一例に過ぎず、本明細書に記載される方法論の実施形態の使用範囲または機能性に関するいかなる制限も示唆することを意図していない。図示された処理システムは、多数の他の汎用または専用コンピューティングシステム環境または構成で動作可能であってもよい。図6に示される処理システムと共に使用するのに適している可能性のある周知のコンピューティングシステム、環境、もしくは構成、またはその組み合わせの例としては、一例として、パーソナルコンピュータシステム、サーバコンピュータシステム、シンクライアント、シッククライアント、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイ

50

クロプロセッサベースのシステム、セットトップボックス、プログラマブル家電、ネットワークPC、ミニコンピュータシステム、メインフレームコンピュータシステム、および上記のシステムまたはデバイスのいずれかを含む分散クラウドコンピューティング環境、および同様のものが挙げられる。

【0041】

コンピュータシステムは、プログラムモジュールなどのコンピュータシステム実行可能命令がコンピュータシステムによって実行されるという一般的な文脈で説明される場合がある。一般に、プログラムモジュールは、特定のタスクを実行する、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、ロジック、データ構造などを含んでもよい。コンピュータシステムは、通信ネットワークを介してリンクされるリモート処理デバイスによってタスクが実行される分散型クラウドコンピューティング環境において実施されてもよい。分散型クラウドコンピューティング環境では、プログラムモジュールは、メモリストレージデバイスを含むローカルおよびリモートコンピュータシステムの記憶媒体の両方に配置されてもよい。

10

【0042】

コンピュータシステムの構成要素は、一例として、1つまたは複数のプロセッサまたは処理ユニット12、システムメモリ16、およびシステムメモリ16を含む様々なシステム構成要素をプロセッサ12に結合するバス14を含んでもよい。プロセッサ12は、本明細書に記載される方法を実行するモジュール30を含んでもよい。モジュール30は、プロセッサ12の集積回路にプログラムされてもよいし、メモリ16、ストレージデバイス18、もしくはネットワーク24またはそれらの組合せからロードされてもよい。

20

【0043】

バス14は、メモリバスまたはメモリコントローラ、周辺バス、加速グラフィックスポート、および様々なバスアーキテクチャのいずれかを使用するプロセッサまたはローカルバスを含む、いくつかのタイプのバス構造のいずれかの1つまたは複数を表すことができる。一例として、かかるアーキテクチャは、インダストリスタンダードアーキテクチャ（ISA）バス、マイクロチャンネルアーキテクチャ（MCA）バス、拡張ISA（EISA）バス、ビデオエレクトロニクススタンダードアソシエーション（VESA）ローカルバス、およびペリフェラルコンポーネントインターコネクト（PCI）バスが含まれる。

【0044】

コンピュータシステムは、様々なコンピュータシステム可読媒体を含んでもよい。そのような媒体は、コンピュータシステムによってアクセス可能な任意の利用可能な媒体であってよく、揮発性媒体と不揮発性媒体、取り外し可能な媒体と取り外し不可能な媒体の両方を含んでもよい。

30

【0045】

システムメモリ16は、ランダムアクセスメモリ（RAM）もしくはキャッシュメモリまたはその両方、またはその他などの揮発性メモリの形態のコンピュータシステム可読媒体を含むことができる。コンピュータシステムは、他の取り外し可能/取り外し不可能な、揮発性/不揮発性のコンピュータシステム記憶媒体をさらに含むことができる。ほんの一例として、ストレージシステム18は、取り外し不可能な不揮発性の磁気媒体（例えば、「ハードドライブ」）からの読み取りおよび磁気媒体への書き込みのために提供され得る。図示しないが、取り外し可能な不揮発性の磁気ディスク（例えば、「フロッピーディスク」）からの読み取りおよび磁気ディスクへの書き込みのための磁気ディスクドライブ、および、CD-ROM、DVD-ROMまたは他の光学媒体などの取り外し可能な不揮発性の光ディスクからの読み出し、または、光ディスクへの書き込みのための光ディスクドライブが、提供され得る。そのような場合、それぞれは、1または複数のデータ媒体インタフェースによってバス14に接続することができる。

40

【0046】

コンピュータシステムは、キーボード、ポインティングデバイス、ディスプレイ28などの1または複数の外部デバイス26、ユーザがコンピュータシステムと対話することを

50

可能にする1または複数のデバイス、もしくはコンピュータシステムが1または複数の他のコンピューティングデバイスと通信することを可能にする任意のデバイス（例えば、ネットワークカード、モデムなど）、またはその組み合わせと通信することもできる。このような通信は、入力/出力（I/O）インタフェース20を介して発生することができる。

【0047】

さらに、コンピュータシステムは、ネットワークアダプタ22を介して、ローカルエリアネットワーク（LAN）、一般的なワイドエリアネットワーク（WAN）、もしくは公衆ネットワーク（例えば、インターネット）、またはその組み合わせなどの1つまたは複数のネットワーク24と通信することが可能である。示されているように、ネットワークアダプタ22は、バス14を介してコンピュータシステムの他の構成要素と通信する。図示されていないが、他のハードウェアコンポーネントもしくはソフトウェアコンポーネントまたはその両方がコンピュータシステムと共に使用され得ることを理解されたい。例としては、マイクロコード、デバイスドライバ、冗長処理装置、外部ディスクドライブアレイ、RAIDシステム、テープドライブ、およびデータアーカイブストレージシステム等が挙げられるが、これらに限定されるものではない。

10

【0048】

本開示はクラウドコンピューティングに関する詳細な説明を含むが、本明細書に記載した教示の実装形態はクラウドコンピューティング環境に限定されない。むしろ、本発明の実装形態は、現在公知のまたは将来開発される他の任意の種類のコピュータ環境と共に実施することができる。クラウドコンピューティングは、設定可能なコンピューティングリソースの共有プール（例えばネットワーク、ネットワーク帯域幅、サーバ、処理、メモリ、記憶装置、アプリケーション、仮想マシンおよびサービス）へ、簡便かつオンデマンドのネットワークアクセスを可能にするためのサービス提供のモデルであり、リソースは、最小限の管理労力または最小限のサービスプロバイダとのやり取りによって速やかに準備（provision）およびリリースできるものである。このクラウドモデルは、少なくとも5つの特性、少なくとも3つのサービスモデル、および少なくとも4つの実装モデルを含むことがある。

20

【0049】

特性は以下の通りである。

30

【0050】

オンデマンド・セルフサービス：クラウドの消費者は、サービスプロバイダとの人的な対話を必要することなく、必要に応じて自動的に、サーバ時間やネットワークストレージなどのコンピューティング能力を一方向的に準備することができる。

【0051】

ブロード・ネットワークアクセス：コンピューティング能力はネットワーク経由で利用可能であり、また、標準的なメカニズムを介してアクセスできる。それにより、異種のシンまたはシッククライアントプラットフォーム（例えば、携帯電話、ラップトップ、PDA）による利用が促進される。

【0052】

リソースプーリング：プロバイダのコンピューティングリソースはプールされ、マルチテナントモデルを利用して複数の消費者に提供される。様々な物理リソースおよび仮想リソースが、需要に応じて動的に割り当ておよび再割り当てされる。一般に消費者は、提供されたリソースの正確な位置を管理または把握していないため、位置非依存（location independence）の感覚がある。ただし消費者は、より高い抽象レベル（例えば、国、州、データセンタ）では場所を特定可能な場合がある。

40

【0053】

迅速な柔軟性（elasticity）：コンピューティング能力は、迅速かつ柔軟に準備することができるため、場合によっては自動的に、直ちにスケールアウトし、また、速やかにリリースされて直ちにスケールインすることができる。消費者にとって、準備に利用可能

50

なコンピューティング能力は無制限に見える場合が多く、任意の時間に任意の数量で購入することができる。

【0054】

測定されるサービス：クラウドシステムは、サービスの種類（例えば、ストレージ、処理、帯域幅、アクティブユーザアカウント）に適したある程度の抽象化レベルでの測定機能を活用して、リソースの使用を自動的に制御し最適化する。リソース使用量を監視、制御、および報告して、利用されるサービスのプロバイダおよび消費者の両方に透明性を提供することができる。

【0055】

サービスモデルは以下の通りである。

【0056】

サービスとしてのソフトウェア（SaaS）：消費者に提供される機能は、クラウドインフラストラクチャ上で動作するプロバイダのアプリケーションを利用できることである。当該そのアプリケーションは、ウェブブラウザ（例えばウェブメール）などのシンクライアントインタフェースを介して、各種のクライアント装置からアクセスできる。消費者は、ネットワーク、サーバ、オペレーティングシステム、ストレージや、個別のアプリケーション機能さえも含めて、基礎となるクラウドインフラストラクチャの管理や制御は行わない。ただし、ユーザ固有の限られたアプリケーション構成の設定はその限りではない。

【0057】

サービスとしてのプラットフォーム（PaaS）：消費者に提供される機能は、プロバイダによってサポートされるプログラム言語およびツールを用いて、消費者が作成または取得したアプリケーションを、クラウドインフラストラクチャに展開（deploy）することである。消費者は、ネットワーク、サーバ、オペレーティングシステム、ストレージを含む、基礎となるクラウドインフラストラクチャの管理や制御は行わないが、展開されたアプリケーションを制御でき、かつ場合によってはそのホスティング環境の構成も制御できる。

【0058】

サービスとしてのインフラストラクチャ（IaaS）：消費者に提供される機能は、オペレーティングシステムやアプリケーションを含む任意のソフトウェアを消費者が展開および実行可能な、プロセッサ、ストレージ、ネットワーク、および他の基本的なコンピューティングリソースを準備することである。消費者は、基礎となるクラウドインフラストラクチャの管理や制御は行わないが、オペレーティングシステム、ストレージ、および展開されたアプリケーションを制御でき、かつ場合によっては一部のネットワークコンポーネント（例えばホストファイアウォール）を部分的に制御できる。

【0059】

展開モデルは以下の通りである。

【0060】

プライベートクラウド：このクラウドインフラストラクチャは、特定の組織専用で運用される。このクラウドインフラストラクチャは、当該組織または第三者によって管理することができる、オンプレミスまたはオフプレミスで存在することができる。

【0061】

コミュニティクラウド：このクラウドインフラストラクチャは、複数の組織によって共有され、共通の関心事（例えば、ミッション、セキュリティ要件、ポリシー、およびコンプライアンス）を持つ特定のコミュニティをサポートする。このクラウドインフラストラクチャは、当該組織または第三者によって管理することができる、オンプレミスまたはオフプレミスで存在することができる。

【0062】

パブリッククラウド：このクラウドインフラストラクチャは、不特定多数の人々や大規模な業界団体に提供され、クラウドサービスを販売する組織によって所有される。

10

20

30

40

50

【0063】

ハイブリッドクラウド：このクラウドインフラストラクチャは、2つ以上のクラウドモデル（プライベート、コミュニティまたはパブリック）を組み合わせたものとなる。それぞれのモデル固有の実体は保持するが、標準または個別の技術によってバインドされ、データとアプリケーションの可搬性（例えば、クラウド間の負荷分散のためのクラウドバースティング）を実現する。

【0064】

クラウドコンピューティング環境は、ステートレス性（statelessness）、低結合性（low coupling）、モジュール性（modularity）および意味論的相互運用性（semantic interoperability）に重点を置いたサービス指向型環境である。クラウドコンピューティングの中核にあるのは、相互接続されたノードのネットワークを含むインフラストラクチャである。

【0065】

図7は、クラウドコンピューティング環境50を示す図である。クラウドコンピューティング環境50は1つまたは複数のクラウドコンピューティングノード10を含む。これらに対して、クラウド消費者が使用するローカルコンピュータ装置（例えば、PDAもしくは携帯電話54A、デスクトップコンピュータ54B、ラップトップコンピュータ54C、もしくは自動車コンピュータシステム54Nまたはこれらの組み合わせなど）は通信を行うことができる。ノード10は互いに通信することができる。ノード10は、例えば、上述のプライベート、コミュニティ、パブリックもしくはハイブリッドクラウドまたはこれらの組み合わせなど、1つまたは複数のネットワークにおいて、物理的または仮想的にグループ化（不図示）することができる。これにより、クラウドコンピューティング環境50は、サービスとしてのインフラストラクチャ、プラットフォームもしくはソフトウェアまたはこれらの組み合わせを提供することができ、クラウド消費者はこれらについて、ローカルコンピュータ装置上にリソースを維持する必要がない。なお、図7に示すコンピュータ装置54A～Nの種類は例示に過ぎず、コンピューティングノード10およびクラウドコンピューティング環境50は、任意の種類ネットワークもしくはネットワークアドレス指定可能接続（例えば、ウェブブラウザの使用）またはその両方を介して、任意の種類電子装置と通信可能であることを理解されたい。

【0066】

ここで図8を参照すると、クラウドコンピューティング環境50（図7）によって提供される一連の機能抽象化レイヤが示されている。なお、図8に示すコンポーネント、レイヤおよび機能は例示に過ぎず、本発明の実施形態はこれらに限定されないことをあらかじめ理解されたい。図示するように、以下のレイヤおよび対応する機能が提供される。

【0067】

ハードウェアおよびソフトウェアレイヤ60は、ハードウェアコンポーネントおよびソフトウェアコンポーネントを含む。ハードウェアコンポーネントの例には、メインフレーム61、縮小命令セットコンピュータ（RISC）アーキテクチャベースのサーバ62、サーバ63、ブレードサーバ64、記憶装置65、ならびにネットワークおよびネットワークコンポーネント66が含まれる。いくつかの実施形態において、ソフトウェアコンポーネントは、ネットワークアプリケーションサーバソフトウェア67およびデータベースソフトウェア68を含む。

【0068】

仮想化レイヤ70は、抽象化レイヤを提供する。当該レイヤから、例えば以下の仮想エンティティを提供することができる：仮想サーバ71、仮想ストレージ72、仮想プライベートネットワークを含む仮想ネットワーク73、仮想アプリケーションおよびオペレーティングシステム74、ならびに仮想クライアント75。

【0069】

一例として、管理レイヤ80は以下の機能を提供することができる。リソース準備81は、クラウドコンピューティング環境内でタスクを実行するために利用されるコンピュー

10

20

30

40

50

ティングリソースおよび他のリソースの動的な調達を可能にする。計量および価格設定 82 は、クラウドコンピューティング環境内でリソースが利用される際のコスト追跡、およびこれらのリソースの消費に対する請求またはインボイス送付を可能にする。一例として、これらのリソースはアプリケーションソフトウェアのライセンスを含んでよい。セキュリティは、データおよび他のリソースに対する保護のみならず、クラウドコンシューマおよびタスクの識別確認を可能にする。ユーザポータル 83 は、コンシューマおよびシステム管理者にクラウドコンピューティング環境へのアクセスを提供する。サービスレベル管理 84 は、要求されたサービスレベルが満たされるように、クラウドコンピューティングリソースの割り当ておよび管理を可能にする。サービス品質保証 (SLA) の計画および履行 85 は、SLA に従って将来必要になると予想されるクラウドコンピューティングリソースの事前手配および調達を可能にする。

10

【0070】

ワークロードレイヤ 90 は、クラウドコンピューティング環境が利用可能な機能の例を提供する。このレイヤから提供可能なワークロードおよび機能の例には、マッピングおよびナビゲーション 91、ソフトウェア開発およびライフサイクル管理 92、仮想教室教育の配信 93、データ分析処理 94、取引処理 95、ならびに、決定木処理 96 が含まれる。

【0071】

本発明は、任意の実行可能な技術的詳細レベルの融合におけるシステム、方法、もしくはコンピュータプログラム製品またはその組み合わせであってよい。コンピュータプログラム製品は、プロセッサに本発明の態様を実行させるためのコンピュータ可読プログラム命令を持つコンピュータ可読記憶媒体 (または複数の媒体) を含んでよい。

20

【0072】

コンピュータ可読記憶媒体は、使用するために命令実行装置によって命令を保持および記憶する有形の装置である。コンピュータ可読記憶媒体は、例えば、これらに限定されないが、電子記憶装置、磁気記憶装置、光記憶装置、電磁記憶装置、半導体記憶装置、または前述を任意に適宜組み合わせたものであってよい。コンピュータ可読記憶媒体のより具体的な一例を列挙すると、ポータブルコンピュータフロッピーディスク、ハードディスク、ランダムアクセスメモリ (RAM)、読み取り専用メモリ (ROM)、書き込みおよび消去可能なランダムアクセスメモリ (EPROM またはフラッシュメモリ)、静的ランダムアクセスメモリ (SRAM)、ポータブルコンパクトディスク読み取り専用メモリ (CD-ROM)、デジタル多用途ディスク (DVD)、メモリスティック、フロッピーディスク、機械的命令が記録された溝を持つパンチカードまたは隆起構造などの符号化されたデバイス、および前述を任意に適宜組み合わせたものである。本明細書で使用されるコンピュータ可読記憶媒体は、電波または他の自由に伝播する電磁波、導波管または他の伝送媒体を通して伝播する電磁波 (例えば、光ファイバーケーブルを通過する光パルス)、またはワイヤーを介して送信される電気信号などの一時的な信号自体であると解釈されるべきではない。

30

【0073】

本明細書に記載のコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体からそれぞれのコンピューティング/処理装置に、または、ネットワーク (例えば、インターネット、ローカルエリアネットワーク、ワイドエリアネットワーク、もしくはワイヤレスネットワークネットワークまたはその組み合わせ) を介して外部コンピュータまたは外部記憶装置にダウンロードすることができる。ネットワークは、銅線伝送ケーブル、光伝送ファイバー、無線伝送、ルーター、ファイアウォール、スイッチ、ゲートウェイコンピュータ、もしくはエッジサーバーまたはその組み合わせで構成される。各コンピューティング/処理装置のネットワークアダプタカードまたはネットワークインターフェースは、ネットワークからコンピュータ可読プログラム命令を受信し、それぞれのコンピューティング/処理装置内のコンピュータ可読記憶媒体に格納するためにコンピュータ可読プログラム命令を転送する。

40

50

【 0 0 7 4 】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セットアーキテクチャ（ISA）命令、機械命令、機械依存命令、マイクロコード、ファームウェア命令、状態設定データ、統合回路のための構成データ、またはSmalltalk、C++などのオブジェクト指向プログラミング言語と「C」プログラミング言語や類似のプログラミング言語などの手続き型プログラミング言語を含む、1つ以上のプログラミング言語の任意の組み合わせで記述されたソースコードまたはオブジェクトコードのいずれかであってよい。コンピュータ可読プログラム命令は、スタンドアロンソフトウェアパッケージとして、完全にユーザのコンピュータ上で、または部分的にユーザのコンピュータ上で実行可能である。あるいは、部分的にユーザのコンピュータ上でかつ部分的にリモートコンピュータ上で、または完全にリモートコンピュータまたはサーバ上で実行可能である。後者のシナリオでは、リモートコンピュータは、ローカルエリアネットワーク（LAN）またはワイドエリアネットワーク（WAN）を含む任意のタイプのネットワークを介してユーザのコンピュータに接続され、または（例えば、インターネットサービスプロバイダを使用したインターネット経由で）外部コンピュータに接続されてよい。いくつかの実施形態では、例えば、プログラマブルロジック回路、フィールドプログラマブルゲートアレイ（FPGA）、またはプログラマブルロジックアレイ（PLA）を含む電子回路は、本発明の態様を実行するために、コンピュータ可読プログラム命令の状態情報を利用してパーソナライズすることにより、コンピュータ可読プログラム命令を実行することができる。

10

20

【 0 0 7 5 】

本発明の態様は、本発明の実施形態による方法、装置（システム）、およびコンピュータプログラム製品のフローチャート図もしくはブロック図またはその両方を参照して本明細書に記載されている。フローチャート図もしくはブロック図またはその両方の各ブロック、およびフローチャート図もしくはブロック図またはその両方のブロックの組み合わせは、コンピュータ可読プログラム命令によって実装できることが理解されよう。

【 0 0 7 6 】

これらのコンピュータ可読プログラム命令は、コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサを介して実行される命令がフローチャートもしくはブロック図またはその両方の1つまたは複数のブロックで指定された機能/動作を実装するための手段を生成するように、機械を生成するためにコンピュータのプロセッサまたは他のプログラム可能なデータ処理装置に提供されることができる。これらのコンピュータ可読記憶媒体はまた、フローチャートもしくはブロック図またはその両方の1つまたは複数のブロックで指定された機能/行為の態様を実装する命令を含む生成品の1つを命令が記憶されたコンピュータ可読プログラム命令が構成するように、コンピュータ、プログラム可能なデータ処理装置、もしくは特定の方法で機能する他のデバイスまたはその組み合わせに接続可能なコンピュータ可読記憶媒体の中に記憶されることができる。

30

【 0 0 7 7 】

コンピュータ、他のプログラム可能な装置、または他のデバイス上でフローチャートもしくはブロック図またはその両方の1つまたは複数のブロックで指定された機能/行為を実行する命令のように、コンピュータ可読プログラム命令はまた、コンピュータ、他のプログラム可能な装置、または他のデバイスにロードされ、コンピュータ、他のプログラム可能な装置、または他のデバイス上で一連の操作ステップを実行し、コンピュータ実装された過程を生成することができる。

40

【 0 0 7 8 】

図中のフローチャートおよびブロック図は、本発明の様々な実施形態によるシステム、方法、およびコンピュータプログラム製品が実行可能な態様の構成、機能、および動作を示している。これに関して、フローチャートまたはブロック図の各ブロックは、モジュール、セグメント、または命令の一部を表してよく、これは、指定された論理機能を実装するための1つまたは複数の実行可能命令を構成する。いくつかの代替の実施形態では、ブ

50

ロックに示されている機能は、図に示されている順序とは異なる場合がある。例えば、連続して示される2つのブロックは、実際には、1つのステップとして達成される場合があり、同時に、実質的に同時に、部分的または全体的に時間的に重複する方法で実行されるか、またはブロックは、機能に応じて逆の順序で実行される場合がある。ブロック図もしくはフローチャート図またはその両方の各ブロック、およびブロック図もしくはフローチャート図またはその両方のブロックの組み合わせは、指定された機能または動作を実行する、または特別な目的のハードウェアとコンピュータ命令の組み合わせを実行する特別な目的のハードウェアベースのシステムによって実装できることにも留意されたい。

【0079】

本明細書で使用される用語は、特定の実施形態を説明することのみを目的としており、本発明を限定することを意図するものではない。本明細書で使用される場合、単数形「a」、「an」および「the」は、文脈が明確に他のことを示さない限り、複数形も含むことを意図している。本明細書で使用される場合、用語「または(or)」は包括的な演算子であり、文脈が明示的にまたは明確に他のことを示さない限り、「および/または(and/or)」を意味することができる。本明細書で使用される場合、用語「含む(comprise)」、「含む(comprises)」、「含む(comprising)」、「含む(include)」、「含む(including)」、もしくは「有する(having)」、またはその組み合わせは、記載された特徴、整数、ステップ、操作、要素、もしくは構成要素またはその組み合わせの存在を指定することができるが、1つ以上の他の特徴、整数、ステップ、操作、要素、コンポーネント、もしくはそれらのグループまたはその組み合わせの存在または追加を排除するものではない。本明細書で使用される場合、「一実施形態において(in an embodiment)」という語句は、同じ実施形態を指す場合もあるが、必ずしも同じ実施形態を指すわけではない。本明細書で使用される場合、「一実施形態において(in one embodiment)」という語句は、同じ実施形態を指す場合もあるが、必ずしも同じ実施形態を指すわけではない。本明細書で使用される場合、「別の実施形態において(in another embodiment)」という語句は、異なる実施形態を指す場合もあるが、必ずしも異なる実施形態を指すわけではない。さらに、実施形態もしくは実施形態の構成要素またはその両方は、相互に排他的でない限り、互いに自由に組み合わせることができる。

【0080】

以下の特許請求の範囲におけるすべての対応する構造、材料、行為、および同等物またはステッププラス関数要素は、特に主張されているように、他の請求された要素と組み合わせる機能を実行するための任意の構造、材料、または行為を含むことを意図している。本発明の説明は、例示および説明の目的で提示されたが、網羅的であることを意図するものではなく、または開示された形態に限定されることを意図するものではない。多くの修正および変形は、本発明の範囲および精神から逸脱することなく、当業者には明らかであろう。実施形態は、本発明の原理および実際の適用を最もよく説明し、当業者が、企図される特定の使用に適した様々な修正を伴う様々な実施形態について本発明を理解できるようにするために選択および説明された。

10

20

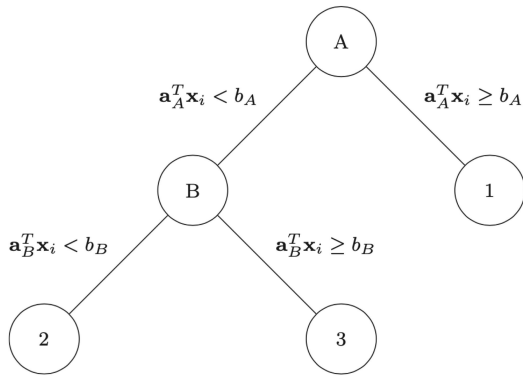
30

40

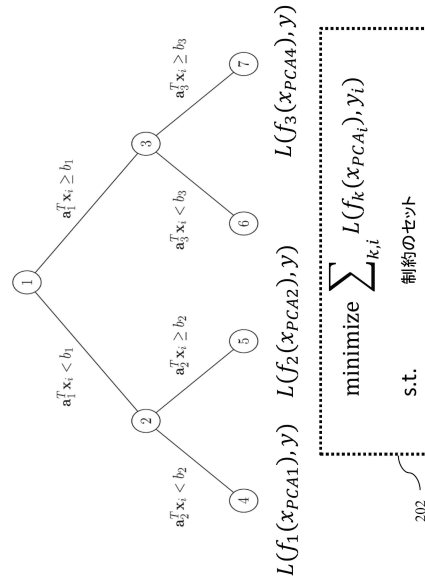
50

【 図面 】

【 図 1 】



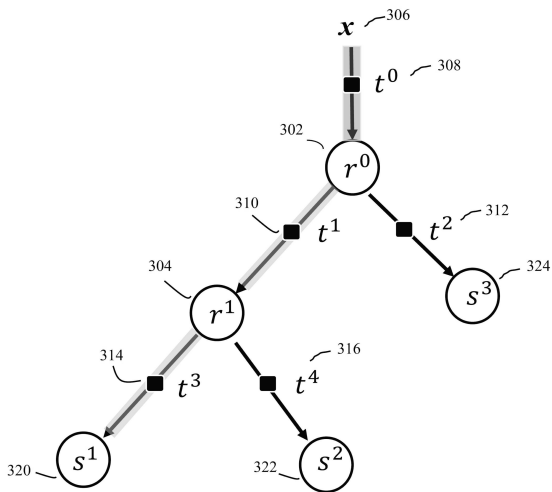
【 図 2 】



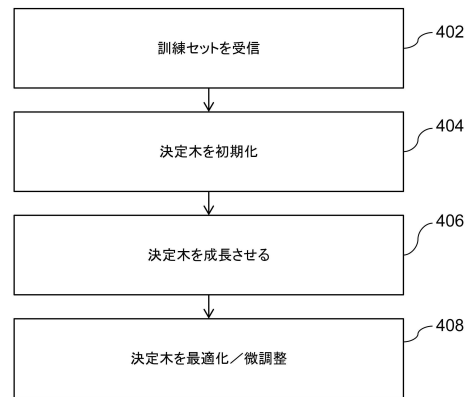
10

20

【 図 3 】



【 図 4 】

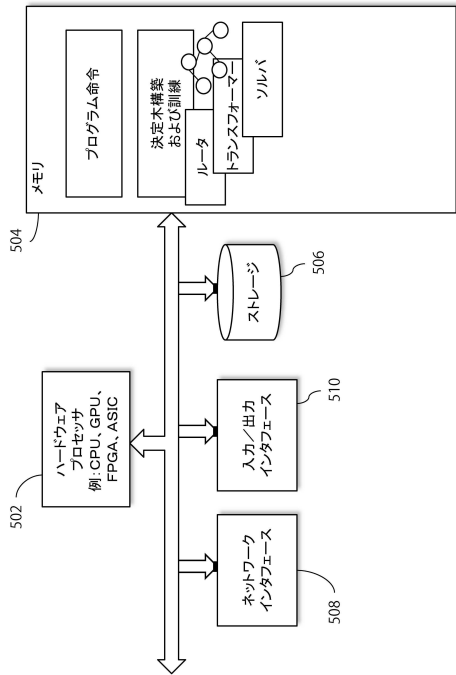


30

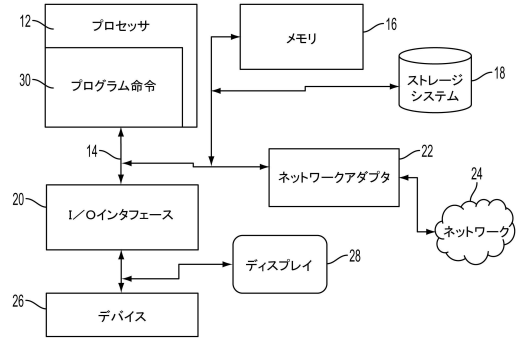
40

50

【 図 5 】



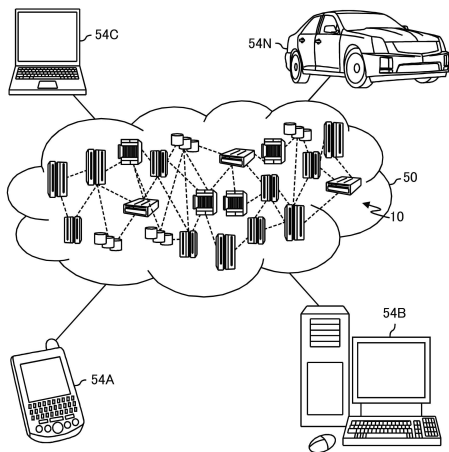
【 図 6 】



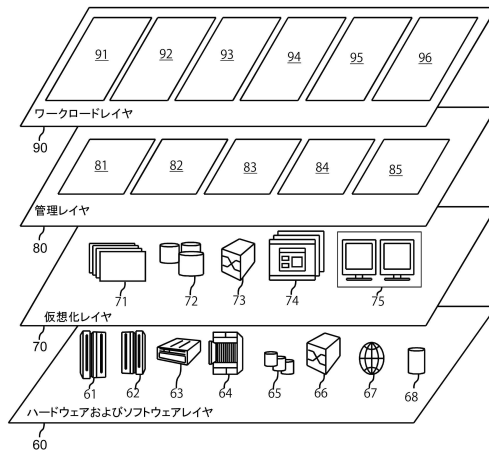
10

20

【 図 7 】



【 図 8 】



30

40

50

フロントページの続き

(54)【発明の名称】 次元削減による分類木と回帰木のパフォーマンス向上)

(74)代理人 100120710

弁理士 片岡 忠彦

(74)復代理人 100104880

弁理士 古部 次郎

(74)復代理人 100118108

弁理士 久保 洋之

(72)発明者 ファン、ティエン、ズン

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード 1 1
0 1 ピーオーボックス2 1 8 インターナショナル・ビジネス・マシーンズ・コーポレーション内

(72)発明者 ファン、マイケル

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード 1 1
0 1 ピーオーボックス2 1 8 インターナショナル・ビジネス・マシーンズ・コーポレーション内

(72)発明者 ムラリ、パヴァンクマール

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード 1 1
0 1 ピーオーボックス2 1 8 インターナショナル・ビジネス・マシーンズ・コーポレーション内

(72)発明者 グエン、ミン、ラム

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード 1 1
0 1 ピーオーボックス2 1 8 インターナショナル・ビジネス・マシーンズ・コーポレーション内