(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第7289086号
(P7289086)
(45)発行日　令和5年6月9日(2023.6.9)　(24)登録日　令和5年6月1日(2023.6.1)

(51)Int.Cl.　　F I
G06N 20/00 (2019.01)　G06N 20/00
G06Q 10/04 (2023.01)　G06Q 10/04

請求項の数 19　(全 24 頁)

(21)出願番号　特願2021-528927(P2021-528927)
(86)(22)出願日　令和1年11月28日(2019.11.28)
(65)公表番号　特表2022-510142(P2022-510142A)
(43)公表日　令和4年1月26日(2022.1.26)
(86)国際出願番号　PCT/IB2019/060288
(87)国際公開番号　WO2020/121104
(87)国際公開日　令和2年6月18日(2020.6.18)
審査請求日　令和4年4月18日(2022.4.18)
(31)優先権主張番号　16/214,703
(32)優先日　平成30年12月10日(2018.12.10)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　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 BUSINESS MACHINES CORPORATION
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, Armonk, New York 10504, 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】インスタンスレベルおよびグループレベルの予測メトリックの事後改善

(57)【特許請求の範囲】
【請求項１】
インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理コンピュータ実装方法であって、
グループ・バイアスへの制約付きで、所定の個人バイアス閾値よりも大きい個人バイアスを有するサンプルを検出することを学習するバイアス検出器を訓練することと、
実行時に、前記所定の個人バイアス閾値よりも大きい個人バイアスを有する実行時サンプル内のバイアスがかかったサンプルを選択するために、前記バイアス検出器を前記実行時サンプルに適用することと、
前記実行時に、前記バイアスがかかったサンプルに対するバイアス除去された予測を提案することと、
を含む、後処理コンピュータ実装方法。
【請求項２】
前記適用することおよび前記提案することは、後処理で動作し、個人バイアスを有する前記バイアスがかかったサンプルを是正の対象として、個人およびグループの両方の公平性メトリックに基づいて前記バイアスがかかったサンプルのバイアスを変更するようにする、請求項１に記載の後処理コンピュータ実装方法。
【請求項３】
前記バイアス検出器は、
前記サンプルを有する訓練セット内の保護属性を摂動させることと、

前記訓練セット内の各サンプル・ポイントについて、複数回の摂動後に平均個人バイアスを求めることと、
有利なクラスに関する前記個人バイアスの差を求めることと、
非特権グループの場合、前記個人バイアスを前記平均個人バイアスとして設定することと、
前記有利なクラスに関する前記個人バイアスの前記差の降順で前記訓練セット内のサンプルをソートすることと、
によって訓練される、請求項１または２に記載の後処理コンピュータ実装方法。
【請求項４】
前記バイアス除去された予測を前記提案することは、
訓練セット内の保護属性を摂動させることと、
前記摂動の結果を顧客モデルにかけることと、
前記顧客モデルにかけられた前記摂動の結果に対する最も尤度の高い予測を採取することと、
この結果が有利なクラスに属する場合、非特権グループ・メンバーに対して前記摂動の結果に対する前記最も尤度の高い予測に変更することと、
によって実行される、請求項１～３のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項５】
前記バイアス検出器によって予測された非特権グループの中から検出された個人バイアス・サンプルが、前記提案することによって優先的に訂正され、
前記提案することは、摂動させたサンプルを顧客モデルにかけ、最も尤度の高い予測を採取することによって、前記バイアスを正すための訂正を実行し、
ユーザは、前記バイアスの元の値、または前記提案されたバイアス除去された予測のいずれを選ぶかを決定する、
請求項１～４のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項６】
前記摂動は保護属性にわたって実行され、集約結果が決定される、請求項５に記載の後処理コンピュータ実装方法。
【請求項７】
複数のクラスの中から結果を選ぶことは、
摂動後の各クラスの集約予測を確認することと、
前記摂動後の最も尤度の高い予測結果を見つけることと、
の一方によって行われる、請求項５または６に記載の後処理コンピュータ実装方法。
【請求項８】
前記訓練中に、前記バイアス検出器は、
非特権グループ・サンプルのペイロード・データ内の保護属性を摂動させ、
前記摂動させたデータと元のデータとの有利な結果の確率の差を求めることにより、非特権グループ・サンプルの個人バイアス・スコアを計算する個人バイアス・チェッカを実装することと、
前記所定の個人バイアス閾値よりも大きい前記個人バイアスを有する前記非特権グループ・サンプルにフラグを立てることと、
前記フラグが立てられたサンプルとフラグが立てられていないサンプルとを区別するように前記バイアス検出器を訓練することと、
によって訓練される、請求項１～７のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項９】
前記実行時に、
前記適用することは、前記実行時に前記バイアス検出器を各非特権グループ・サンプルに適用して、前記個人バイアスの尤度を計算し、

保護属性を摂動させ、摂動後の結果をチェックすることにより、前記個人バイアスがかかったサンプルをさらにテストし、
前記摂動後の前記結果が元の結果と異なる場合、個人バイアスがかかったサンプルをバイアス除去された予測としてアービタに提案し、前記アービタは、前記バイアスの元の値と、バイアス除去された予測とから選ぶことができる、
請求項１～８のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項１０】
クラウド・コンピューティング環境で具現化される、請求項１～９のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項１１】
インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理コンピュータ・プログラムであって、前記コンピュータ・プログラムは、コンピュータに、
グループ・バイアスへの制約付きで、所定の個人バイアス閾値よりも大きい個人バイアスを有するサンプルを検出することを学習するバイアス検出器を訓練することと、
実行時に、前記所定の個人バイアス閾値よりも大きい個人バイアスを有する実行時サンプル内のバイアスがかかったサンプルを選択するために、前記バイアス検出器を前記実行時サンプルに適用することと、
前記実行時に、前記バイアスがかかったサンプルに対するバイアス除去された予測を提案することと、
を実行させる、後処理コンピュータ・プログラム。
【請求項１２】
インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理システムであって、
プロセッサと、
メモリと、
を備え、前記メモリは前記プロセッサに、
グループ・バイアスへの制約付きで、所定の個人バイアス閾値よりも大きい個人バイアスを有するサンプルを検出することを学習するバイアス検出器を訓練することと、
実行時に、前記所定の個人バイアス閾値よりも大きい個人バイアスを有する実行時サンプル内のバイアスがかかったサンプルを選択するために、前記バイアス検出器を前記実行時サンプルに適用することと、
前記実行時に、前記バイアスがかかったサンプルに対するバイアス除去された予測を提案することと、
を実行させるための命令を記憶する、後処理システム。
【請求項１３】
インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理コンピュータ実装方法であって、
所定の個人閾値よりも大きい値を有する実行時サンプル内のサンプルを選択するために、訓練された検出器を前記実行時サンプルに適用することと、
前記サンプルに対する変更された予測を提案することと、
を含む、後処理コンピュータ実装方法。
【請求項１４】
前記適用することおよび前記提案することは、個人バイアスを有する前記サンプルを是正の対象として、個人およびグループの両方の公平性メトリックに基づいて前記サンプルの前記値を変更するようにする後処理で動作する、請求項１３に記載の後処理コンピュータ実装方法。
【請求項１５】
前記訓練された検出器は、完全なブラックボックスとして受け取られる、請求項１３または１４に記載の後処理コンピュータ実装方法。

【請求項１６】
前記適用することは、検証セットにグラウンド・トゥルース・クラス・ラベルを必要としない純粋な実行時の手法を含む、請求項１３～１５のいずれか１項に記載の後処理コンピュータ実装方法。
【請求項１７】
インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理コンピュータ実装方法であって、
個人バイアスについて、非特権グループの各サンプルをテストすることと、
前記サンプルが所定の個人バイアス閾値より大きいバイアスを有する場合、特権グループの結果を前記サンプルに割り当てることと、
を含む、後処理コンピュータ実装方法。
【請求項１８】
前記非特権グループは、前記非特権グループの分類に応じて前記個人バイアスに基づいて前記結果を受け取る、請求項１７に記載の後処理コンピュータ実装方法。
【請求項１９】
前記特権グループ内のサンプルの前記結果は、前記割り当てることによって変更されない、請求項１７または１８に記載の後処理コンピュータ実装方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、一般に後処理方法に関連し、より詳細には、限定としてではなく、差別的効果（ｄｉｓｐａｒａｔｅ　ｉｍｐａｃｔ）のグループ公平性尺度を改善することを目的としたバイアス軽減アルゴリズムにおいてデータサンプルに優先順位を付けるために使用される個人バイアス検出器を介して後処理で個人およびグループの両方の公平性を高めるためのシステム、方法、および記録媒体に関する。
【０００２】
従来、バイアスがかかった分類器の予測の公平性を高めるための後処理手法は、グループの公平性のみに対処している。公平性、無差別、および望ましくないバイアスは、人間の意思決定において常に懸念事項であったが、過去の人間の意思決定が、雇用、信用、および刑事司法などの危険度の高い用途における機械学習モデル用の訓練データとして現在使用されているので、ますます注目されている。バイアスを軽減しなければ、そのような意思決定に関して訓練されたモデルは、人間のバイアスを永続させ、増減させるので、安全でなく、信頼できない。最近では、検出、推定理論、および情報理論を使用して基本原理が定義された機械学習およびデータ・マイニングの文献では、アルゴリズムの公平性に関する活動が活発になっている。
【０００３】
従来、意思決定における公平性には、グループの公平性および個人の公平性という２つの主要な概念がある。グループの公平性は、その最も広い意味において、母集団を保護属性によって定義されたグループに分割し、ある統計的尺度がグループ間で等しくなることを求める。異なる統計的尺度に関する多くの異なるグループの公平性の概念が存在し、そのような概念の１つは差別的効果である。個人の公平性は、その最も広い意味において、類似の個人が同様に扱われることを求める。グループの公平性のチェックは、統計的メトリックの簡単な計算であるが、個人の公平性のチェックは、多くの値を有する多くの保護属性が存在する場合により計算が複雑になり、モデルを使用してサンプルをスコアリングするのにはコストがかかる。最近では、不平等指数に基づくグループおよび個人の両方の公平性に関する統一されたメトリックが提案されている。
【０００４】
機械学習パイプラインには、望ましくないバイアスを軽減するための３つの可能な介入ポイント、すなわち、訓練データ、学習手順、および出力予測が含まれ、３つの対応するクラスの、すなわち、前処理、処理中、および後処理のバイアス軽減アルゴリズムがある

。後処理アルゴリズムの利点は、訓練プロセスへのアクセスを必要としないので、実行時環境に適していることである。さらに、後処理アルゴリズムはブラックボックス方式で動作し、すなわち、モデルの内部、その派生物（ｄｅｒｉｖａｔｉｖｅｓ）などにアクセスする必要がないので、任意の機械学習モデルに適用可能である。

【０００５】

従来、バイアス軽減アルゴリズムの大多数はグループの公平性に対処しているが、個人の公平性に対処しているものは少数である。一部の前処理アルゴリズムは、グループおよび個人の両方の公平性に対処している。既存の後処理アルゴリズムは全て、グループの公平性のためだけのものである。

【０００６】

したがって、グループおよび個人の両方の公平性を考慮した後処理バイアス軽減技法が当技術分野で必要とされている。さらに、バイアス軽減アルゴリズムの訓練中に、検証サンプルにグラウンド・トゥルース・クラス・ラベルを必要としない技法が当技術分野で必要とされている。

【０００７】

したがって、効果的な説明方法を生成するように構成することができる攻略法が当技術分野で必要とされている。

【発明の概要】

【０００８】

後処理アルゴリズムの一般的な方法論は、サンプルのサブセットを取得し、グループ公平性要件を満たすように、予測されたラベルを適切に変更することである。後処理に関する興味深い観察は、メトリックが期待値であるので、グループ公平性要件を達成するように任意のサンプルを改変できることである。当技術分野における問題を考慮して、本発明者らは、個人の公平性の問題を有するまたは有し得るサンプルを選び、これにより、グループおよび個人の両方の公平性に一緒に対処することができる。当技術分野における問題を解決する本発明の手法の出発点は、個人バイアス検出器であり、これは、保護属性を変化させ、他の全ての特徴を一定のままにした場合に、モデル予測が変化するサンプルを見つけるものである。この技法で多くの効率化が実現されるが、それでも計算コストは高くなる。検出器を継続的に実行できないという制限を克服するために、本発明は、小さなポイント・セットで個人の公平性をチェックし、新しいサンプルに適用される分類器を訓練することによって、それらから汎化する。個人バイアスを有し得るサンプルは、予測されたラベルの変更が検討されるサンプルである。そうすることにより、本発明は、不確定性に焦点を当てることから個人バイアスに焦点を当てることへと当技術分野を改善する。

【０００９】

例示的な実施形態では、本発明は、インスタンスレベルおよびグループレベルの予測メトリックの事後改善のための後処理コンピュータ実装方法であって、グループ・バイアスへの制約付きで、所定の個人バイアス閾値よりも大きい個人バイアスを有するサンプルを検出することを学習するバイアス検出器を訓練することと、所定の個人バイアス閾値よりも大きいバイアスを有する実行時サンプル内のバイアスがかかったサンプルを選択するために、バイアス検出器を実行時サンプルに適用することと、バイアスがかかったサンプルに対するバイアス除去された予測を提案することと、を含む、後処理コンピュータ実装方法を提供することができる。

【００１０】

１つまたは複数の他の例示的な実施形態は、コンピュータ・プログラム製品およびシステムを含む。

【００１１】

当技術分野への本貢献をよりよく理解できるように、本発明の他の詳細および実施形態を以下に説明する。それにもかかわらず、本発明は、その適用において、説明に記載しているかまたは図面に示しているそのような詳細、表現、用語、図解、もしくは配置またはそれらの組み合わせに限定されない。むしろ、本発明は、記載したものに加えた実施形態

が可能であり、様々な方法で実施および実行することが可能であり、限定的なものと見なされるべきではない。

【００１２】

したがって、当業者は、本開示の基礎となる概念が、本発明のいくつかの目的を実行するための他の構造、方法、およびシステムの設計の基礎として容易に利用され得ることを理解するであろう。そのため、特許請求の範囲が、本発明の主旨および範囲から逸脱しない限り、そのような同等の構造を含むものと見なされることが重要である。

【００１３】

本発明の態様は、図面を参照して、本発明の例示的な実施形態の以下の詳細な説明からよりよく理解されよう。

【図面の簡単な説明】

【００１４】

【図１】後処理方法１００の高レベルのフロー・チャートを例示的に示す図である。

【図２】本発明の一実施形態による方法１００の訓練フェーズを例示的に示す図である。

【図３】本発明の一実施形態による方法１００の実行時フェーズを例示的に示す図である。

【図４】本発明の一実施形態による方法１００の第１のアルゴリズムを例示的に示す図である。

【図５】本発明の一実施形態による方法１００の第１の結果を例示的に示す図である。

【図６】本発明の一実施形態による方法１００の第２の結果を例示的に示す図である。

【図７】本発明の一実施形態による方法１００の第３の結果を例示的に示す図である。

【図８】本発明の一実施形態による方法１００の第４の結果を例示的に示す図である。

【図９】本発明の一実施形態によるクラウド・コンピューティング・ノード１０を示す図である。

【図１０】本発明の一実施形態によるクラウド・コンピューティング環境５０を示す図である。

【図１１】本発明の一実施形態による抽象化モデル・レイヤを示す図である。

【発明を実施するための形態】

【００１５】

【数１】

$$\hat{y}$$

【００１６】

は以降ｙハットと記載する。

【００１７】

【数２】

$$\check{y}$$

【００１８】

は以降ｙキャロン（ｃａｒｏｎ）と記載する。

【００１９】

【数３】

$$\dot{b}$$

【００２０】

は以降ｂドットと記載する。

【００２１】

以下、図１～図１１を参照して本発明を説明し、図全体を通して、同様の参照番号は同様の部分を指す。一般的な慣行に準じて、図面の様々な特徴は必ずしも縮尺通りではないことを強調しておく。それどころか、様々な特徴の寸法は、明確にするために任意に拡大

または縮小する場合がある。

【００２２】

ここで図１に示す例を参照すると、後処理方法１００は、任意の確率的機械学習モデル（またはモデルの融合）に適用可能なブラックボックス手法である、実行時環境に適した様々なステップを含み、また、個人およびグループの両方の公平性メトリックを改善する。

【００２３】

図９に示すように、本発明の一実施形態によるコンピュータ・システム１２の１つまたは複数のコンピュータは、図１のステップを実行するための命令がストレージ・システムに記憶されたメモリ２８を含むことができる。

【００２４】

１つまたは複数の実施形態（たとえば、図９～図１１を参照）は、クラウド環境５０（たとえば、図１０を参照）に実装され得るが、それにもかかわらず、本発明はクラウド環境以外に実装できることは理解されよう。

【００２５】

図１～図８全体に関連して、本発明では、特徴$X \in X$、カテゴリ的保護属性$D \in D$、およびカテゴリ・ラベル$Y \in Y$を有する教師あり分類問題を考える。訓練サンプルのセット$\{(x_1, d_1, y_1), \ldots, (x_n, d_n, y_n)\}$が与えられると、本発明は分類器（すなわち、ステップ１０１）yハット：$X \times D \rightarrow Y$を学習する。

【００２６】

「分類器」および「バイアス検出器」は同じ意味で使用していることに留意されたい。

【００２７】

説明を簡単にするために、スカラーのバイナリ保護属性（すなわち、$D = \{0, 1\}$）と、バイナリ分類問題（すなわち、$Y = \{0, 1\}$）とを考える。値$d = 1$は、「特権グループ」（たとえば、刑事司法適用における米国の第１のグループ）に対応するように設定され、$d = 0$は、「非特権グループ」（たとえば、第１のグループと比較して否定的な扱いを受ける、またはビジネスの場で融資を受けられたり受けられなかったりする、刑事司法適用における米国の第２のグループ）に対応するように設定される。値$y = 1$は「有利な結果」（たとえば、融資を受けられたり、逮捕されたりしないこと）に対応するように設定され、$y = 0$は「不利な結果」（たとえば、融資を受けられなかったり、逮捕されたりすること）に対応するように設定される。

【００２８】

コンテキストに基づいて、連続出力スコアyハット$_s \in [0, 1]$を有する確率的バイナリ分類器も、$\{0, 1\}$への閾値と共に使用される。一方、本発明は、複数の保護属性およびマルチクラスの結果の場合に拡張可能である。実際に、複数のクラスが、複数の「有利」または「不利」な結果を有し得る。

【００２９】

個人バイアスの定義の１つは次のとおりである。yハット$(x_i, d = 0) \neq$ yハット$(x_i, d = 1)$の場合、サンプルiは個人バイアスを有する。$b_i = I[$yハット$(x_i, d = 0) \neq$ yハット$(x_i, d = 1)]$とし、ここで、$I[\cdot]$は指示関数である。個人バイアス・スコア$b_{s,i} =$ yハット$_s(x_i, d = 1) -$ yハット$_s(x_i, d = 0)$は、b_iのソフト・バージョンである。個人バイアス要約統計量を計算するために、テスト・サンプルにわたってb_iの平均が取られる。

【００３０】

「差別的効果」として知られているグループの公平性の１つの概念は、次のように定義される。式（４）が$1 - \varepsilon$未満であるか、または$(1 - \varepsilon) - 1$より大きい場合、「差別的効果」が存在し、ここで、εの一般的な値は０．２である。

【００３１】

【数４】

$$\frac{\mathbb{E}[\hat{y}(\mathbf{X}, D) \mid D = 0]}{\mathbb{E}[\hat{y}(\mathbf{X}, D) \mid D = 1]} \qquad (4)$$

【００３２】

個人バイアス検出用のテスト生成に関して、個人バイアス検出における２つの特有の問題がある。１つ目は、個人バイアスのケースの有無を判定することである。２つ目は、全サンプルの個人バイアス・ステータスを決定することである。

【００３３】

本発明はこれに対し、１つ目の問題に関して、任意のブラック・ボックス分類器（たとえば、顧客モデル）の決定空間を体系的に探索して、バイアスがかかっている可能性が高いテスト・サンプルを生成する技法の改良を含む。

【００３４】

この方法は、２種類の検索を使用する。１つ目はグローバル検索であり、これにより、様々な領域が網羅されるように決定空間が探索される。２つ目はローカル検索であり、これにより、発見済みの個人バイアスがかかったサンプルの非保護特徴の値を巧みに摂動（ｐｅｒｔｕｒｂ）させることによって、テスト・ケースが生成される。重要なアイデアの一例は、動的シンボリック実行、すなわち、プログラム・パス内の制約を無効化して検索制約を生成し、制約ソルバを使用して新しい検索パスを見つける、プログラム用の既存の体系的なテスト・ケース生成技法を使用することである。このアルゴリズムは、多数の属性および属性値を含む状況でサンプルのバッチに使用された場合に、計算の観点から２つ目の特有の問題を解決するのに役立つ。

【００３５】

しかしながら、許容可能なグループの公平性を達成するために、分類器ｙハット$_i$のラベル出力を他のラベルｙキャロン$_i \in$Ｙに変更する様々な後処理方法が適用され得る。棄却オプション分類（ＲＯＣ：ｒｅｊｅｃｔ　ｏｐｔｉｏｎ　ｃｌａｓｓｉｆｉｃａｔｉｏｎ）法は、あるマージン・パラメータθについて、｜ｙハット$_s$－０．５｜＜θとなる不確定サンプルを考え（０．５が分類閾値であると仮定する）、ｄ$_i$＝０のサンプルにはｙキャロン$_i$＝１を割り当て、ｄ$_i$＝１のサンプルにはｙキャロン$_i$＝０を割り当てる。いわゆる「棄却オプション」バンド外の特定のサンプルについては、ｙキャロン$_i$＝ｙハット$_i$である。θの値は、差別的効果に関する要件を達成するように最適化され得る。

【００３６】

本発明は、公平性後処理アルゴリズムを含む。差別的効果のようなグループ公平性メトリックに関する後処理における重要な観察は、それらが期待値として定義されているので、個人のサンプルが交換可能であることである。本発明は、個人バイアスを有し得るＸの部分からサンプルを選出する。本発明はこれを、まず個人バイアス検出を行い（たとえば、訓練されたバイアス検出器を使用）、次に後処理バイアス軽減アルゴリズムを使用する（たとえば、非特権グループの結果を有利なクラスの結果に変更する）ことによって行う。

【００３７】

個人バイアス検出について、訓練データセット・パーティションで訓練済みの分類器ｙハットを考える。本発明は、上記で提供した個人バイアスの定義を、それに従うラベルを有さない検証パーティションで評価することができる。これらの検証サンプルには個人バイアスを有するものと有さないものがある。Ｘにおける個人バイアスのある程度の一貫性または平滑性の仮定の下で、本発明は、（たとえば、ステップ１０１において）この検証セットから個人バイアスの分類器または検出器（たとえば、バイアス検出器）を学習することができ、これは個人バイアスが未知である未見のサンプルへと汎化する。

【００３８】

ここで、スコア出力を提供する任意の分類または異常検出アルゴリズムを使用し得る。

本発明では例示的に、経験的結果にロジスティック回帰を使用する。

【００３９】

正式には、非特権グループ（$d_j = 0$）に属する検証セット・サンプル（x_j，d_j）、$j = 1, \ldots, m$のd_jを摂動させることによって、本発明は個人バイアス・スコア$b_{s,j}$を求める。方法１００はさらに、データセット｛（x_1，β_1），．．．，（x_m，β_m）｝を構築し、これを使用して個人バイアス検出器bドット（・）を訓練する。最大の個人バイアス・スコアを有するサンプルの場合、β_jは１であり、残りの場合は０である。この割り当ては、検証セット全体に対する差別的効果の制約に基づいて選ばれた個人バイアス・スコアに対する閾値τ（たとえば、「所定の個人バイアス閾値」）によって決定される。これは、差別的効果の要件に基づいてマージン・パラメータが調整されるアルゴリズムに似ている。

【００４０】

たとえば、図２は、バイアス検出器の訓練段階を例示的に示している。訓練では、ペイロード・データ２５０（たとえば、サンプル）を受け取り、これらは、摂動２４０を使用する（すなわち、検証セット・サンプルのd_jを摂動させる）個人バイアス・チェッカ２０１にかけられて、個人バイアス・スコアが求められる。チェッカ２０１は、結果に対する制約である顧客モデル２２０と協議する（たとえば、誰が承認されるべきか、または承認されないべきか）。顧客モデルは、見込み客（銀行、政府、企業など）から受け取った例示的なブラックボックス・データ・セットである。データセット｛（x_1，β_1），．．．，（x_m，β_m）｝を使用して、汎化された個人バイアス・チェッカ２０１からの個人バイアス・スコアの入力を用いて、バイアス・サンプル検出器２０２を訓練する。すなわち、個人バイアス・スコアは、摂動２４０によって訓練セット内の保護属性（複数可）を摂動させ、各サンプル・ポイントについて、場合によっては複数回の摂動後に平均スコアを求め、有利なクラスに関するスコアの差を求め、ここで、特権グループの場合、元のスコアおよび平均摂動後スコアを求め、「非特権グループ」の場合、元のスコアに対する平均摂動後スコアを求め、スコア差の降順で（個人バイアスの高いものから低いものへ）サンプルをソートすることによって求められる。

【００４１】

上述のように、グループ公平性制約２３０を使用してバイアス・サンプル検出器２０２を訓練し、ここで、最大の個人バイアス・スコアを有するサンプルの場合、β_jは１であり、残りの場合は０である。この割り当ては、検証セット全体に対する差別的効果の制約に基づいて選ばれた個人バイアス・スコアに対する閾値τ（たとえば、「所定の個人バイアス閾値」）によって決定される。訓練されたバイアス・サンプル検出器２０２は、ソートされたサンプルからバイアスがかかったサンプル・ポイントの最適なセットを識別し、上から開始して各ポイントをバイアス除去し、グループ・バイアスをチェックすることを許容レベルに達するまで行い、バイアス除去された標本（ｅｘａｍｐｌｅ）に＋１のラベルを付け、バイアス除去されなかった標本に「－１」のラベルを付ける。本発明では、非特権グループの標本のみにラベルを付ける。特権グループの標本は考慮しない。これらのラベル付けされた標本から、バイアスがかかったサンプルを検出する分類器（最近傍、ロジスティック回帰、ランダム・フォレストなど）が訓練される。この分類器は、ソフト・バイアス尤度を提供可能でなければならない。持続検出器（ｐｅｒｓｉｓｔ　ｄｅｔｅｃｔｏｒ）２６０は、バイアス検出器モデルを将来の使用のためにメモリに保存する。

【００４２】

これにより、訓練の際に、任意のサンプルを同等に改変することにより、グループ公平性メトリックを改善できるので、個人バイアスを有し得るサンプルに焦点を当てることができる。これは、計算コストの高い個人バイアス・チェッカから汎化して、個人バイアスを有し得る新しいサンプルを識別するモデルを作成し、まず、これらのサンプルを改変してグループ公平性メトリック要件を達成するようにする。

【００４３】

なお、訓練された個人バイアス検出器は不要であるという主張もあり得、全てのサンプ

ルについて、実行時に来たときに単にb_iを計算すればよく、その理由は、そうするためには、ブラックボックス分類器モデルを使用したスコアリングしか必要ないためである。これは真実であり得るが、以下の注意点がある。第一に、d_iがスカラーかつバイナリであることが仮定されており、多くの場合はそうではない。そのため、b_iを計算するには複数回のモデル評価が必要になり得、これは特に、スコアリングされるサンプルごとに、モデルを配備してバイアスを是正するエンティティによって支払われる一定の金額がかかることが想定される産業用途では、法外になり得る。第二に、本発明は、グループ公平性制約に基づいてバイナリのβ_j値を計算し、これによって、最大の個人バイアス・スコアを有する標本のみに対してバイアス除去が検討され、過剰な支払いがなくなる。全ての標本が$b_i=1$で等しくバイアスがかかっていると見なされる場合、このレベルの制御は不可能である。

【００４４】

個人バイアス検出器ｂドットが検証セットで訓練されると（すなわち、ステップ１０１）、実行時にバイアス軽減アルゴリズムが適用されて、次のようにサンプルがテストされる（すなわち、ステップ１０２）。非特権グループ（$d_i=0$）からの各サンプルは、個人バイアスのテストを受け、個人バイアスを有し得る場合（すなわち、$b_{ドット i}=1$）、このサンプルには、それが有利なクラスにあれば受け取ったはずの結果が割り当てられる（すなわち、$y_{キャロン i}=y_{ハット}(x_{k,\,l})$）。発明と同様の人間の感受性をエンコードするために、特権グループからのサンプルを含め、他の全てのサンプルは未変更のままにする。図４に示すアルゴリズムは、上記の説明を例示的にまとめたものである。

【００４５】

図３に示すように、実行時環境（すなわち、ステップ１０２～１０３）には、訓練済みの顧客モデルに加え、本発明のバイアス検出器モデルが既に存在する。定期的な間隔で、個人バイアス・チェックがテスト・サンプルに対して実行され、特徴空間全体に汎化される。新しいラベルのないサンプルが入ってくると、モデルはそれをスコアリングし、汎化された個人バイアス・チェッカは、それが個人バイアスを有するか否かを予測する。これらの予測は後処理バイアス軽減アルゴリズムに供給され、これは、個人およびグループの両方の公平性メトリックを達成するように（たとえば、摂動および提案３０３と、提案されたバイアス除去された予測３５０とを介して）スコアを改変する方法を決定する。最後のアービタ３４０は、元のモデル予測または改変された予測を最終出力予測３６０として選ぶオプションをユーザに与える。

【００４６】

すなわち、実行時のバイアス・サンプル検出器およびバイアス除去は、テスト標本３３０に対して、バイアス検出器３０１の適用を通じて、それらのソフト・バイアス尤度に従って降順で適用される。個人バイアス尤度が閾値よりも大きい場合（３０２で判定される）、本発明はそれらの標本について３０３以降を続ける。本発明では、グループ・バイアスをチェックしない。このため、提案手法は、実行時に１つの標本にさえ適用することができる。

【００４７】

これは、「非特権グループ」の標本に対してのみ実行される。３０２で識別されたサンプルから、上から開始して各ポイントをバイアス除去し、グループ・バイアスをチェックすることを、許容レベルに達するまで行う。バイアス除去された予測３５０は後処理で、各サンプル・ポイントに対するバイアス除去手順によって実行され、この手順は、訓練セット内の保護属性（複数可）を摂動させ（たとえば、図３の３０３）、摂動させた標本を顧客モデル３２０にかけ、摂動させたデータに対する最も尤度の高い予測を修正すべき提案値として採取することによって行われる。

【００４８】

これにより、検出器によって（「非特権グループ」の中で）最大の個人バイアスを有すると予測されたサンプルが優先的に訂正され、提案される訂正は、摂動させた標本を顧客モデルにかけ、最も尤度の高い予測を採取することを含み、アービタは、元の予測または

提案されたバイアス除去された予測のいずれを選択するかを決定することができる。また、本発明は非特権グループのみに焦点を当てているので、特権グループについては、本発明は個人バイアスを全く計算しない。

【００４９】

一実施形態では、摂動が保護属性にわたって実行され、集約スコア／結果が決定されることを除いて、動作は上記と同様である。他の実施形態では、複数のクラスの場合、本発明がバイナリ・クラスではなく複数のクラスの中から結果を選ぶことを除いて、動作は同様である。これはいくつかの方法で、すなわち、摂動後の各クラスの集約予測スコアを確認し、摂動後の最も尤度の高い予測結果を見つけること、または任意のユーザ定義の方法で、実行することができる。「クラス」は「特権または非特権」であり得る。

【００５０】

アルゴリズムの公平性はビジネスおよび社会にとって重要なトピックであり、できるだけ多くの公平性の側面に対処する新しいバイアス軽減アルゴリズムを開発することは重要である。本明細書に開示する本発明は、個人およびグループの両方の公平性メトリックを改善するために、個人バイアスを有するサンプルを是正の対象とする新しい後処理アルゴリズムを含み、実験によりいくつかの実世界のデータセットで分類精度をそれほど損なうことなくそうすることを示した。機械学習業界は、モデル構築とモデル配備とが分離されるパラダイムに向かって進んでいる。これには、配備者が訓練済みのモデルの内部にアクセスするための能力が限られていることが含まれる。そのため、後処理アルゴリズム、特に分類器を完全なブラックボックスとして扱うことができるアルゴリズムが必要である。以前の技術と比較して、方法１００は、個人およびグループの両方の公平性に取り組むだけでなく、検証セットにグラウンド・トゥルース・クラス・ラベルを必要としないので、純粋な実行時の手法にもなる。

【００５１】

「クラス」は通常、結果を表すことに留意されたい。たとえば、融資が承認された場合は「有利」と見なされ、融資が拒否された場合は「不利」と見なされる。「グループ」は、保護されたグループなどの人口統計群を表す。たとえば、１つ目の人口統計群を「特権グループ」とすることができ、２つ目の人口統計群を「非特権グループ」とすることができる。

【００５２】

本発明の上記の説明を考慮すると、本発明は、機械学習モデルからの予測に関するインスタンスレベルおよびグループレベルの両方のメトリックを事後的に改善し得る。グループレベルのメトリックの一例は、グループ・バイアスである。グループレベルおよびインスタンスレベルのメトリックは、たとえば、ある程度の効用を各個人に提供するシステムを含むことができ、ユーザは、適格な個人に対するあるレベルの保証に加え、集約されたグループ・レベルでのある程度の保証を確保することを望む。この場合、本発明は提案手法の変形を使用することができる。一般的に、グループ・レベルでの平均的な保証によって、適格な個人ごとの保証は確保されない。

【００５３】

本発明は、グループ・バイアスへの制約付きで、高い個人バイアスを有するサンプルを検出することを学習するバイアス検出器を訓練することによって、インスタンスレベルおよびグループレベルのメトリックを改善する。次に、本発明は、バイアス検出器を新規の実行時サンプルに適用して、バイアスを有し得るサンプルを選び、それらに対するバイアス除去された予測を提案する。

【００５４】

バイアス検出器の訓練段階の間に、本発明は、非特権グループ・サンプルのペイロード・データの保護属性を摂動させ、摂動させたデータと元のデータとの有利な結果の確率の差を求めることによって、それらの個人バイアス・スコアを計算する個人バイアス・チェッカを実装する。複数回の摂動の場合、有利な結果の平均確率が使用される。高い個人バイアスを有する非特権グループ・サンプルには「バイアスがかかったサンプル」のフラグ

が立てられ、グループ・バイアス制約が満たされるまでこの処理が繰り返される。バイアス検出器は、フラグが立てられたサンプルとフラグが立てられていないサンプルとを区別するように訓練される（たとえば、図２を参照）。

【００５５】

実行時／適用段階の間に、本発明は、実行時に各非特権グループ・サンプルにバイアス検出器を適用し、個人バイアスの尤度を計算する。最も尤度の高いバイアスがかかったサンプルが、さらなるテストのために選ばれる。さらなるテストは、保護属性を摂動させ、摂動後の予測をチェックすることを含む。複数回の摂動の場合、複数回の摂動の結果が単一の結果に集約される。摂動後の結果が元の結果と異なる場合、これはバイアス除去された予測としてアービタに提案され、アービタは、さらなる考慮事項に基づいて、元の予測と、バイアス除去された予測とから選ぶことができる（たとえば、図３を参照）。

【００５６】

本発明は、保護属性に対して複数回の摂動を実行することができる。このため、本発明は、数値的保護属性およびカテゴリ的保護属性の両方を扱うことができる。複数の保護属性を使用することもできる。また、本発明は、選択された結果が有利と見なされ、残りが不利と見なされる、一般的なカテゴリ的な結果を扱うことができる。

【００５７】

実験結果

上記の方法１００は、ＵＣＩ　Ａｄｕｌｔ（１９９４年の米国国勢調査データベースに基づく収入データセット；４５，２２２サンプル；有利な結果：＄５０，０００を超える収入；保護属性：性別、人種）、ＵＣＩ　Ｓｔａｔｌｏｇ　Ｇｅｒｍａｎ　Ｃｒｅｄｉｔ（クレジット・スコアリング・データセット；１，０００サンプル；有利な結果：低リスク；保護属性：性別、年齢）、およびＰｒｏＰｕｂｌｉｃａ　ＣＯＭＰＡＳ（刑務所の常習性データセット；６，１６７サンプル；有利な結果：再犯しない；保護属性：性別、人種）の３つの標準データセットで評価する。３つのデータセットのそれぞれは、２つのバイナリ保護属性を有し、我々はこれらを２つの異なる問題と見なすので、全部で６つの問題が生じる。方法１００の結果は、棄却オプション分類（ＲＯＣ）技法、および等化オッズ後処理（ＥＯＰ：ｅｑｕａｌｉｚｅｄ　ｏｄｄｓ　ｐｏｓｔ－ｐｒｏｃｅｓｓｉｎｇ）技法と比較される。

【００５８】

これらのデータセットは、例示的なＡＩ　Ｆａｉｒｎｅｓｓ　３６０ツールキットを使用して処理およびロードされ、６０％の訓練パーティション、２０％の検証パーティション、および２０％のテスト・パーティションにランダムに分割される。これらの実験は、データセットの２５個のそのようなランダムなパーティションを用いて実行され、以下の実験結果にエラー・バーを設けることが可能になる。訓練パーティションを使用して、実験では、ブラックボックス分類器として、L_2正則化ロジスティック回帰と、ランダム・フォレストとの両方を組み込む。ランダム・フォレストの場合、実験では、ツリーの数を１００個に設定し、リーフ・ノードあたりの最小サンプル数を２０個に設定する。

【００５９】

３つのバイアス軽減手法全てのパラメータは、データセットの検証パーティションで最適化される。ＲＯＣおよびＥＯＰの両方の手法では、検証セットにグラウンド・トゥルース・クラス・ラベルが必要であるが、純粋な実行時の方法である方法１００では必要ない。ＲＯＣおよび方法１００は、範囲（０．８，１．２５）で、すなわち、$\varepsilon = 0.2$で、「差別的効果」を達成するように最適化される。「差別的効果」ではなく等化オッズ用に設計されたＥＯＰは、差別的効果の範囲に対して最適化することはできない。

【００６０】

以下のサブセクションでは、実験は最初に、提案方法１００で使用される個人バイアス検出器の有効性を実証し、その後、分類精度、差別的効果、および個人の公平性について３つのアルゴリズムを比較する。

【００６１】

図６～図８において、「ＩＧＤ」ラベルは方法１００の結果に対応することに留意されたい。

【００６２】

個人バイアスの汎化に関する検証結果の図を提供する図５を参照すると、実験では、未見のテスト・データに対する個人バイアス検出器の汎化性能を検証する。個人バイアス検出器は、非特権グループ・サンプル（ｄ＝０）のみで使用されるので、その性能尺度はこのサブセットのみについて計算される。バイアス検出器のグラウンド・トゥルース・ラベルは、テスト・データ内の全ての「非特権グループ」のサンプルに対して個人バイアス・スコア（$b_{s,k}$）を実際に計算し、「差別的効果」の制約に基づいてグラウンド・トゥルース・バイアス・ラベル（β_k）を識別することによって、得られる。これらのラベルは、バイアス検出器（$b\hat{}_k$）によって予測されたラベルと比較され、バランス分類精度が計算される。

【００６３】

バイアス検出器のこの性能を、ブラックボックス分類器がロジスティック回帰である場合の全てのデータセットおよび保護属性の組み合わせについて、図５に示す。全ての精度値は０．８５を超えており、これは目下の目的に対する明確な有効性を示している。検出器は、ブラックボックス分類器がランダム・フォレストの場合も同様に機能し、最小精度は約０．８０である。

【００６４】

公平性の比較のために、実験では、ＥＯＰ、ＲＯＣ、および方法１００を比較するために、（ａ）個人バイアス、（ｂ）差別的効果、および（ｃ）バランス分類精度、の３つの尺度を使用する。これらの尺度は、後処理された予測ｙキャロンを使用して計算される。個人バイアス尺度は、上記の段落００３８で説明した要約統計量であり、差別的効果尺度は式（４）で定義され、バランス分類精度は、真のラベルｙに対する予測ｙキャロンに関して求められた真陽性率および真陰性率の平均である。実験では、元の（オリジナルの）予測ｙハットに関するこれらの尺度も求める。図６～図８に示すように、方法１００は、元の分類器に近い精度を維持しながら、両方の公平性尺度を一貫して改善する唯一の方法である。全ての結果は、ブラックボックス分類器としてのロジスティック回帰に関して示しているが、ランダム・フォレストでも同様の結果が観察される（スペースの制約のため省略している）。

【００６５】

実験に基づいて、個人バイアスでは、方法１００はＧｅｒｍａｎデータセットおよびＣＯＭＰＡＳデータセットで最高性能を発揮している。ＲＯＣ法は、Ａｄｕｌｔデータセットに対して最高性能を発揮しているが、バランス精度を低下させるという犠牲を払っている。ＥＯＰ法およびＲＯＣ法は、個人バイアスを増加させることがあるが、方法１００には全く当てはまらない。方法１００はまた、元の予測に対して差別的効果を一貫して改善しており、ただし、６つのケースのうち５つでＲＯＣ法が上回っている。ＲＯＣ手法の強力な性能は、個人バイアスも最適化していないためと考えられる。ＥＯＰ法は、差別的効果に関して性能が低く、これはおそらく、オッズを等化するように設計されていたために、「差別的効果」が改善される場合と、常に改善されない場合とがあるためである。

【００６６】

また、方法１００は、検証パーティションでグラウンド・トゥルース・ラベルが使用されていないにもかかわらず、元の予測と比較してバランス分類器精度を維持するのに最適である。

【００６７】

クラウド・コンピューティング環境を使用した例示的な態様

この詳細な説明はクラウド・コンピューティング環境における本発明の例示的な実施形態を含むが、本明細書に列挙した教示の実装形態はそのようなクラウド・コンピューティング環境に限定されないことを理解されたい。むしろ、本発明の実施形態は、現在知られているまたは今後開発される他の任意のタイプのコンピューティング環境と共に実装する

ことが可能である。

【００６８】

クラウド・コンピューティングは、最小限の管理労力またはサービスのプロバイダとのやり取りによって迅速に供給および公開することができる、設定可能なコンピューティング・リソース（たとえば、ネットワーク、ネットワーク帯域幅、サーバ、処理、メモリ、ストレージ、アプリケーション、仮想マシン、およびサービス）の共有プールへの便利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る。このクラウド・モデルは、少なくとも５つの特徴と、少なくとも３つのサービス・モデルと、少なくとも４つの配備モデルとを含み得る。

【００６９】

特徴は以下の通りである。

オンデマンド・セルフ・サービス：クラウド・コンシューマは、サービスのプロバイダとの人的な対話を必要とせずに、必要に応じて自動的に、サーバ時間およびネットワーク・ストレージなどのコンピューティング能力を一方的に供給することができる。

ブロード・ネットワーク・アクセス：能力はネットワークを介して利用することができ、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（たとえば、携帯電話、ラップトップ、およびＰＤＡ）による使用を促進する標準的なメカニズムを介してアクセスされる。

リソース・プーリング：プロバイダのコンピューティング・リソースをプールして、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マルチ・テナント・モデルを使用して複数のコンシューマにサービス提供する。一般にコンシューマは、提供されるリソースの正確な位置に対して何もできず、知っているわけでもないが、より高い抽象化レベル（たとえば、国、州、またはデータセンターなど）における位置を特定可能であり得るという点で位置非依存の感覚がある。

迅速な弾力性：能力を迅速かつ弾力的に、場合によっては自動的に供給して素早くスケール・アウトし、迅速に解放して素早くスケール・インすることができる。コンシューマにとって、供給可能な能力は無制限であるように見えることが多く、任意の時間に任意の数量で購入することができる。

測定されるサービス：クラウド・システムは、サービスのタイプ（たとえば、ストレージ、処理、帯域幅、およびアクティブ・ユーザ・アカウント）に適したある抽象化レベルでの計量機能を活用して、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コンシューマの両方に透明性を提供することができる。

【００７０】

サービス・モデルは以下の通りである。

ソフトウェア・アズ・ア・サービス（ＳａａＳ：Ｓｏｆｔｗａｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ）：コンシューマに提供される能力は、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使用することである。アプリケーションは、ウェブ・ブラウザ（たとえば、ウェブベースの電子メール）などのシン・クライアント・インターフェースを介して様々なクライアント回路からアクセス可能である。コンシューマは、限定されたユーザ固有のアプリケーション構成設定を可能性のある例外として、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、さらには個々のアプリケーション機能を含む、基盤となるクラウド・インフラストラクチャを管理も制御もしない。

プラットフォーム・アズ・ア・サービス（ＰａａＳ：Ｐｌａｔｆｏｒｍ ａｓ ａ Ｓｅｒｖｉｃｅ）：コンシューマに提供される能力は、プロバイダによってサポートされるプログラミング言語およびツールを使用して作成された、コンシューマ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上に配備することである。コンシューマは、ネットワーク、サーバ、オペレーティング・システム、またはストレージを含む、基盤となるクラウド・インフラストラクチャを管理も制御もしないが、配備されたアプリケーションおよび場合によってはアプリケーション・ホスティング環境構成を制御

する。
インフラストラクチャ・アズ・ア・サービス（ＩａａＳ：Ｉｎｆｒａｓｔｒｕｃｔｕｒｅ　ａｓ　ａ　Ｓｅｒｖｉｃｅ）：コンシューマに提供される能力は、オペレーティング・システムおよびアプリケーションを含むことができる任意のソフトウェアをコンシューマが配備して動作させることが可能な、処理、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・リソースを供給することである。コンシューマは、基盤となるクラウド・インフラストラクチャを管理も制御もしないが、オペレーティング・システム、ストレージ、配備されたアプリケーションを制御し、場合によっては選択したネットワーキング・コンポーネント（たとえば、ホスト・ファイアウォール）を限定的に制御する。

【００７１】
配備モデルは以下の通りである。
プライベート・クラウド：クラウド・インフラストラクチャは組織専用に運用される。これは組織または第三者によって管理され得、構内または構外に存在し得る。
コミュニティ・クラウド：クラウド・インフラストラクチャはいくつかの組織によって共有され、共通の懸念（たとえば、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の考慮事項など）を有する特定のコミュニティをサポートする。これは組織または第三者によって管理され得、構内または構外に存在し得る。
パブリック・クラウド：クラウド・インフラストラクチャは、一般大衆または大規模な業界団体に対して利用可能にされ、クラウド・サービスを販売する組織によって所有される。
ハイブリッド・クラウド：クラウド・インフラストラクチャは、固有のエンティティのままであるが、データおよびアプリケーションの移植性を可能にする標準化技術または独自技術（たとえば、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）によって結合された２つ以上のクラウド（プライベート、コミュニティ、またはパブリック）を合成したものである。

【００７２】
クラウド・コンピューティング環境は、ステートレス性、低結合性、モジュール性、および意味論的相互運用性に重点を置いたサービス指向型である。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。

【００７３】
ここで図９を参照すると、クラウド・コンピューティング・ノードの一例の概略図が示されている。クラウド・コンピューティング・ノード１０は、適切なノードの一例に過ぎず、本明細書に記載の本発明の実施形態の使用または機能性の範囲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。いずれにしても、クラウド・コンピューティング・ノード１０は、本明細書に記載の機能性のいずれかを実装されること、もしくは実行すること、またはその両方が可能である。

【００７４】
クラウド・コンピューティング・ノード１０はコンピュータ・システム／サーバ１２として図示しているが、他の多くの汎用または専用の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ることを理解されたい。コンピュータ・システム／サーバ１２での使用に適し得るよく知られているコンピューティング・システム、環境、もしくは構成、またはそれらの組み合わせの例には、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、ハンドヘルドもしくはラップトップ回路、マルチプロセッサ・システム、マイ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、プログラム可能な家庭用電化製品、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もしくは回路のいずれか含む分散型クラウド・コンピューティング環境などが含まれるが、これらに限定されない。

【００７５】
コンピュータ・システム／サーバ１２は、コンピュータ・システムによって実行されるプログラム・モジュールなどのコンピュータ・システム実行可能命令の一般的なコンテキストで記述され得る。一般に、プログラム・モジュールには、特定のタスクを実行するか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、データ構造などが含まれ得る。コンピュータ・システム／サーバ１２は、通信ネットワークを介してリンクされたリモート処理回路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で実施され得る。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・ストレージ回路を含むローカルおよびリモート両方のコンピュータ・システム記憶媒体に配置され得る。
【００７６】
図９を再度参照すると、コンピュータ・システム／サーバ１２は、汎用コンピューティング回路の形態で示されている。コンピュータ・システム／サーバ１２のコンポーネントは、１つまたは複数のプロセッサまたは処理ユニット１６と、システム・メモリ２８と、システム・メモリ２８を含む様々なシステム・コンポーネントをプロセッサ１６に結合するバス１８と、を含み得るが、これらに限定されない。
【００７７】
バス１８は、メモリ・バスまたはメモリ・コントローラ、ペリフェラル・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うちのいずれかの１つまたは複数を表す。限定ではなく例として、そのようなアーキテクチャには、業界標準アーキテクチャ（ＩＳＡ：Ｉｎｄｕｓｔｒｙ Ｓｔａｎｄａｒｄ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、マイクロ・チャネル・アーキテクチャ（ＭＣＡ：Ｍｉｃｒｏ Ｃｈａｎｎｅｌ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、拡張ＩＳＡ（ＥＩＳＡ：Ｅｎｈａｎｃｅｄ ＩＳＡ）バス、ビデオ・エレクトロニクス規格協会（ＶＥＳＡ：Ｖｉｄｅｏ Ｅｌｅｃｔｒｏｎｉｃｓ Ｓｔａｎｄａｒｄｓ Ａｓｓｏｃｉａｔｉｏｎ）ローカル・バス、および周辺機器相互接続（ＰＣＩ：Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔｓ）バスが含まれる。
【００７８】
コンピュータ・システム／サーバ１２は、典型的には、様々なコンピュータ・システム可読媒体を含む。そのような媒体は、コンピュータ・システム／サーバ１２によってアクセス可能な任意の利用可能な媒体であり得、揮発性および不揮発性の媒体、取り外し可能および取り外し不可能な媒体の両方を含む。
【００７９】
システム・メモリ２８は、ランダム・アクセス・メモリ（ＲＡＭ：ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ）３０もしくはキャッシュ・メモリ３２またはその両方などの、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／サーバ１２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み得る。単なる例として、取り外し不可能な不揮発性の磁気媒体（図示せず、典型的には「ハード・ドライブ」と呼ばれるもの）に読み書きするためのストレージ・システム３４を設けることができる。図示していないが、取り外し可能な不揮発性の磁気ディスク（たとえば、「フレキシブル・ディスク」）に読み書きするための磁気ディスク・ドライブと、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学ディスクに読み書きするための光学ディスク・ドライブと、を設けることができる。そのような例では、それぞれを、１つまたは複数のデータ・メディア・インターフェースによってバス１８に接続することができる。以下でさらに図示および説明するように、メモリ２８は、本発明の実施形態の機能を実行するように構成されるプログラム・モジュールのセット（たとえば、少なくとも１つ）を有する少なくとも１つのプログラム製品を含み得る。
【００８０】

プログラム・モジュール４２のセット（少なくとも１つ）を有するプログラム／ユーティリティ４０は、限定ではなく例として、オペレーティング・システム、１つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データと同様に、メモリ２８に記憶され得る。オペレーティング・システム、１つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データまたはそれらの何らかの組み合わせのそれぞれは、ネットワーク環境の実装形態を含み得る。プログラム・モジュール４２は、一般に、本明細書に記載の本発明の実施形態の機能もしくは方法論またはその両方を実行する。

【００８１】

コンピュータ・システム／サーバ１２はまた、キーボード、ポインティング回路、ディスプレイ２４などの１つまたは複数の外部デバイス１４、ユーザがコンピュータ・システム／サーバ１２とやりとりすることを可能にする１つまたは複数の回路、ならびに／あるいはコンピュータ・システム／サーバ１２が１つまたは複数の他のコンピューティング回路と通信することを可能にする任意の回路（たとえば、ネットワーク・カード、モデムなど）と通信し得る。そのような通信は、入力／出力（Ｉ／Ｏ：Ｉｎｐｕｔ／Ｏｕｔｐｕｔ）インターフェース２２を介して行うことができる。またさらに、コンピュータ・システム／サーバ１２は、ネットワーク・アダプタ２０を介して、ローカル・エリア・ネットワーク（ＬＡＮ：ｌｏｃａｌ　ａｒｅａ　ｎｅｔｗｏｒｋ）、一般的なワイド・エリア・ネットワーク（ＷＡＮ：ｗｉｄｅ　ａｒｅａ　ｎｅｔｗｏｒｋ）、もしくはパブリック・ネットワーク（たとえば、インターネット）、またはそれらの組み合わせなどの、１つまたは複数のネットワークと通信することができる。図示のように、ネットワーク・アダプタ２０は、バス１８を介してコンピュータ・システム／サーバ１２の他のコンポーネントと通信する。図示していないが、他のハードウェアもしくはソフトウェアまたはその両方のコンポーネントを、コンピュータ・システム／サーバ１２と併用できることを理解されたい。例には、マイクロコード、回路ドライバ、冗長処理ユニット、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・システムなどが含まれるが、これらに限定されない。

【００８２】

ここで図１０を参照すると、例示的なクラウド・コンピューティング環境５０が示されている。図示のように、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を含み、これらを使用して、たとえば、パーソナル・デジタル・アシスタント（ＰＤＡ：ｐｅｒｓｏｎａｌ　ｄｉｇｉｔａｌ　ａｓｓｉｓｔａｎｔ）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、または自動車コンピュータ・システム５４Ｎあるいはそれらの組み合わせなどの、クラウド・コンシューマによって使用されるローカル・コンピューティング回路が通信し得る。ノード１０は互いと通信し得る。これらは、たとえば、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、またはそれらの組み合わせなどの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され得る（図示せず）。これにより、クラウド・コンピューティング環境５０は、クラウド・コンシューマがローカル・コンピューティング回路上にリソースを維持する必要がない、インフラストラクチャ・アズ・ア・サービス、プラットフォーム・アズ・ア・サービス、またはソフトウェア・アズ・ア・サービス、あるいはそれらの組み合わせを提供することが可能になる。図１０に示すコンピューティング回路５４Ａ～Ｎのタイプは例示的なものにすぎないことが意図されており、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、任意のタイプのネットワークまたはネットワーク・アドレス指定可能接続（たとえば、Ｗｅｂブラウザを使用）あるいはその両方を介して任意のタイプのコンピュータ化回路と通信できることを理解されたい。

【００８３】

ここで図１１を参照すると、クラウド・コンピューティング環境５０（図１０）によって提供される機能的抽象化レイヤの例示的なセットが示されている。図１１に示すコンポ

ーネント、レイヤ、および機能は例示的なものにすぎないことが意図されており、本発明の実施形態はこれらに限定されないことを事前に理解されたい。図示のように、以下のレイヤおよび対応する機能が提供される。

【００８４】

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、ＲＩＳＣ（縮小命令セット・コンピュータ：Ｒｅｄｕｃｅｄ　Ｉｎｓｔｒｕｃｔｉｏｎ　Ｓｅｔ　Ｃｏｍｐｕｔｅｒ）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、ストレージ回路６５、ならびにネットワークおよびネットワーキング・コンポーネント６６が含まれる。一部の実施形態では、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。

【００８５】

仮想化レイヤ７０は、仮想エンティティの以下の例：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の提供元になり得る抽象化レイヤを提供する。

【００８６】

一例では、管理レイヤ８０は、下記の機能を提供し得る。リソース供給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ティング・リソースおよび他のリソースの動的調達を提供する。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ときのコスト追跡と、これらのリソースの消費に対する課金または請求とを提供する。一例では、これらのリソースはアプリケーション・ソフトウェア・ライセンスを含み得る。セキュリティは、クラウド・コンシューマおよびタスクの識別情報検証だけでなく、データおよび他のリソースに対する保護も提供する。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な、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提供する。サービス・レベル合意（ＳＬＡ）の計画および履行８５は、ＳＬＡに従って将来要求され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提供する。

【００８７】

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され得る機能性の例を提供する。このレイヤから提供され得るワークロードおよび機能の例は、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、仮想教室教育配信９３、データ分析処理９４、取引処理９５、および、より詳細には、本発明に関連して、方法１００、を含む。

【００８８】

本発明は、任意の可能な技術的詳細レベルの統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れらの組み合わせであり得る。コンピュータ・プログラム製品は、本発明の態様をプロセッサに実行させるためのコンピュータ可読プログラム命令をその上に有するコンピュータ可読記憶媒体（または複数の媒体）を含み得る。

【００８９】

コンピュータ可読記憶媒体は、命令実行デバイスによる使用のために命令を保持および記憶できる有形のデバイスとすることができる。コンピュータ可読記憶媒体は、たとえば、限定はしないが、電子ストレージ・デバイス、磁気ストレージ・デバイス、光学ストレージ・デバイス、電磁ストレージ・デバイス、半導体ストレージ・デバイス、またはこれらの任意の適切な組み合わせであり得る。コンピュータ可読記憶媒体のより具体的な例の非網羅的なリストには、ポータブル・コンピュータ・ディスケット、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：ｒｅａｄ－ｏｎ

ｌｙ ｍｅｍｏｒｙ）、消去可能プログラム可能読み取り専用メモリ（ＥＰＲＯＭ：ｅｒａｓａｂｌｅ ｐｒｏｇｒａｍｍａｂｌｅ ｒｅａｄ－ｏｎｌｙ ｍｅｍｏｒｙまたはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ：ｓｔａｔｉｃ ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ）、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：ｐｏｒｔａｂｌｅ ｃｏｍｐａｃｔ ｄｉｓｃ ｒｅａｄ－ｏｎｌｙ ｍｅｍｏｒｙ）、デジタル多用途ディスク（ＤＶＤ：ｄｉｇｉｔａｌ ｖｅｒｓａｔｉｌｅ ｄｉｓｋ）、メモリー・スティック（登録商標）、フレキシブル・ディスク、命令が記録されたパンチ・カードまたは溝の隆起構造などの機械的にコード化されたデバイス、およびこれらの任意の適切な組み合わせが含まれる。コンピュータ可読記憶媒体は、本明細書で使用する場合、たとえば、電波または他の自由に伝搬する電磁波、導波管もしくは他の伝送媒体を伝搬する電磁波（たとえば、光ファイバ・ケーブルを通過する光パルス）、または有線で伝送される電気信号など、一過性の信号自体であると解釈されるべきではない。

【００９０】

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／処理デバイスに、あるいは、たとえば、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、もしくは無線ネットワーク、またはそれらの組み合わせなどのネットワークを介して外部コンピュータまたは外部ストレージ・デバイス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、無線伝送、ルータ、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、またはそれらの組み合わせを含み得る。各コンピューティング／処理デバイスのネットワーク・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、コンピュータ可読プログラム命令を転送して、それぞれのコンピューティング／処理デバイス内のコンピュータ可読記憶媒体に記憶する。

【００９１】

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：ｉｎｓｔｒｕｃｔｉｏｎ－ｓｅｔ－ａｒｃｈｉｔｅｃｔｕｒｅ）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、集積回路の構成データ、あるいは、Ｓｍａｌｌｔａｌｋ（登録商標）、Ｃ＋＋などのオブジェクト指向プログラミング言語、および「Ｃ」プログラミング言語または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つまたは複数のプログラミング言語の任意の組み合わせで書かれたソース・コードまたはオブジェクト・コードであり得る。コンピュータ可読プログラム命令は、完全にユーザのコンピュータ上で、部分的にユーザのコンピュータ上で、スタンドアロン・ソフトウェア・パッケージとして、部分的にユーザのコンピュータ上かつ部分的にリモート・コンピュータ上で、あるいは完全にリモート・コンピュータまたはサーバ上で実行し得る。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得、または（たとえば、インターネット・サービス・プロバイダを使用してインターネットを介して）外部コンピュータに接続され得る。一部の実施形態では、たとえば、プログラマブル論理回路、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ：ｆｉｅｌｄ－ｐｒｏｇｒａｍｍａｂｌｅ ｇａｔｅ ａｒｒａｙ）、またはプログラマブル・ロジック・アレイ（ＰＬＡ：ｐｒｏｇｒａｍｍａｂｌｅ ｌｏｇｉｃ ａｒｒａｙ）を含む電子回路は、本発明の態様を実行するために、電子回路を個人向けにするための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、コンピュータ可読プログラム命令を実行し得る。

【００９２】

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

明細書で説明している。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におけるブロックの組み合わせが、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は理解されよう。

【００９３】

これらのコンピュータ可読プログラム命令を、汎用コンピュータ、専用コンピュータ、または他のプログラム可能データ処理装置のプロセッサに提供して、コンピュータまたは他のプログラム可能データ処理装置のプロセッサを介して実行されたそれらの命令が、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において指定された機能／行為を実装するための手段を生成するようなマシンを生成し得る。また、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可読記憶媒体に記憶され得、それによって、コンピュータ、プログラム可能データ処理装置、もしくは他のデバイス、またはそれらの組み合わせに特定の方法で機能するように指示することができ、その結果、命令が記憶されたコンピュータ可読記憶媒体は、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において指定された機能／行為の態様を実装する命令を含む製造品を含む。

【００９４】

また、コンピュータ可読プログラム命令をコンピュータ、他のプログラマブルデータ処理装置、または他のデバイスにロードして、コンピュータ、他のプログラマブル装置、または他のデバイス上で一連の動作ステップを実行させることによって、コンピュータ、他のプログラマブル装置、または他のデバイス上で実行されたそれらの命令が、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において指定された機能／行為を実装するようなコンピュータ実装処理を生成し得る。

【００９５】

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・プログラム製品の可能な実装形態のアーキテクチャ、機能性、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、指定された論理的機能（複数可）を実装するための１つまたは複数の実行可能命令を含むモジュール、セグメント、または命令の一部を表し得る。一部の代替的実装形態では、ブロックに示す機能は、図に示す順序以外で行われ得る。たとえば、関与する機能性に応じて、連続して示す２つのブロックは、実際には実質的に同時に実行され得、またはそれらのブロックは、場合により逆の順序で実行され得る。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におけるブロックの組み合わせは、指定された機能もしくは行為を実行する専用のハードウェア・ベースのシステムによって実装されるか、または専用ハードウェアおよびコンピュータ命令の組み合わせを実行することができることにも留意されたい。

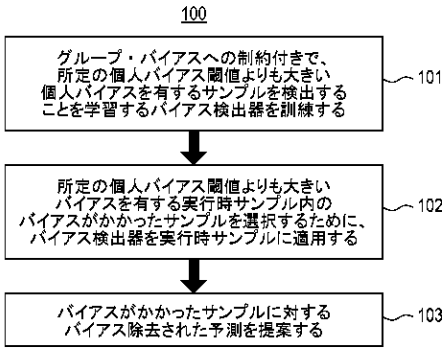
【００９６】

本発明の様々な実施形態の説明は、例示の目的で提示しているが、網羅的であることも、開示した実施形態に限定されることも意図したものではない。開示した実施形態の範囲および主旨から逸脱することなく、多くの修正および変形が当業者には明らかであろう。本明細書で使用している用語は、実施形態の原理、市場で見られる技術に対する実際の適用または技術的改善を最もよく説明するために、または当業者が本明細書に開示した実施形態を理解できるようにするために選択している。

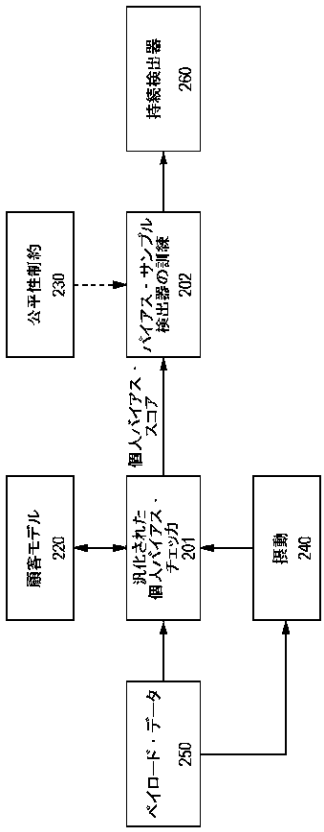
【００９７】

さらに、出願人の意図は、全ての請求項の要素の均等物を包含することであり、本出願のいかなる請求項への補正も、補正した請求項の任意の要素または特徴の均等物に対する利益または権利の放棄として解釈されるべきではない。

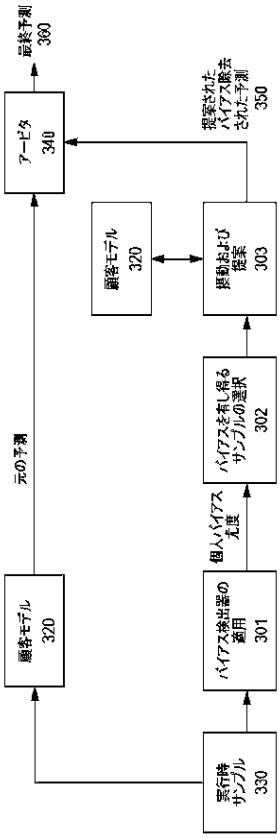
【図１】



【図２】



【図３】

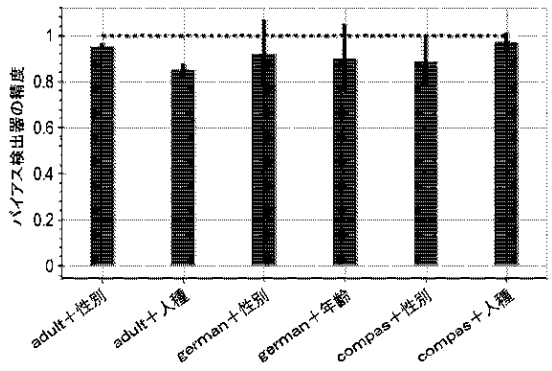


【図４】

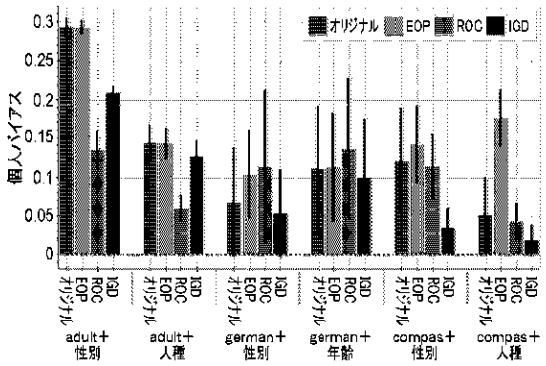
Algorithm 1 Individual+Group Debiasing (IGD) Post-Processing

Given classifier $\hat{y}$ trained on training set $\{(\mathbf{x}_i, d_i, y_i)\}$, and
Given validation set $\{\mathbf{x}_j \mid d_j = 0\}$, compute individual bias scores $\{b_{S,j} \mid d_j = 0\}$.
if $b_{S,j} > \tau$ **then**
  $\beta_j \leftarrow 1$
else
  $\beta_j \leftarrow 0$
end if
Construct auxiliary dataset $\{(\mathbf{x}_j, \beta_j) \mid d_j = 0\}$.
Train individual bias detector $\hat{b}$ on auxiliary dataset.
for all run-time test samples $(\mathbf{x}_k, d_k)$ **do**
  $\hat{y}_k \leftarrow \hat{y}(\mathbf{x}_k, d_k)$
  if $d_k == 0$ **then**
    $\hat{b}_k \leftarrow \hat{b}(\mathbf{x}_k)$
    if $\hat{b}_k == 1$ **then**
      $\tilde{y}_k \leftarrow \hat{y}(\mathbf{x}_k, 1)$
    else
      $\tilde{y}_k \leftarrow \hat{y}_k$
    end if
  else
    $\tilde{y}_k \leftarrow \hat{y}_k$
  end if
end for

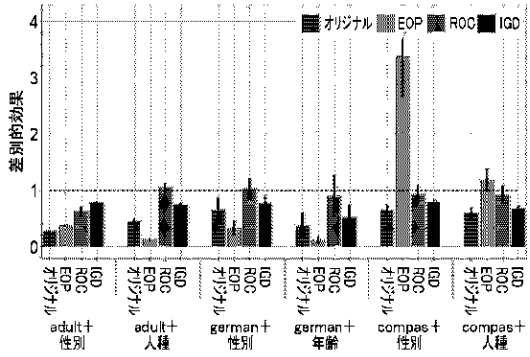
【図５】



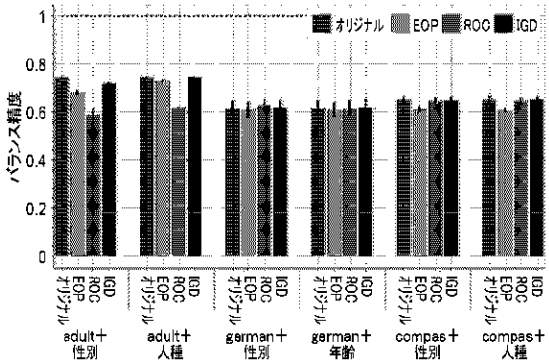
【図６】



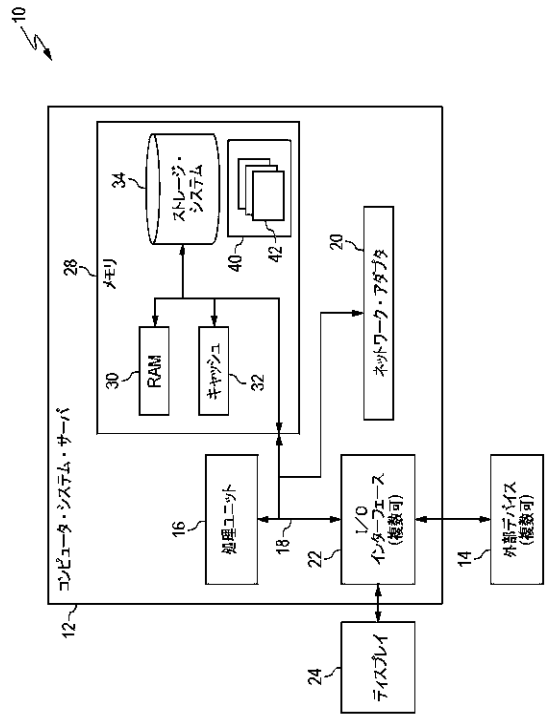
【図７】



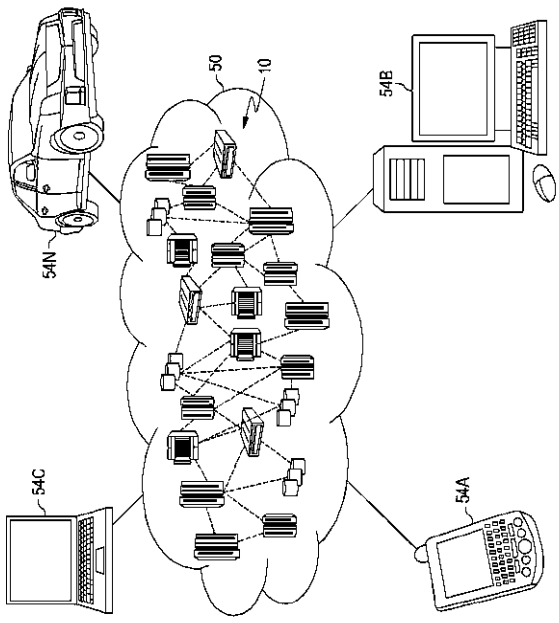
【図８】



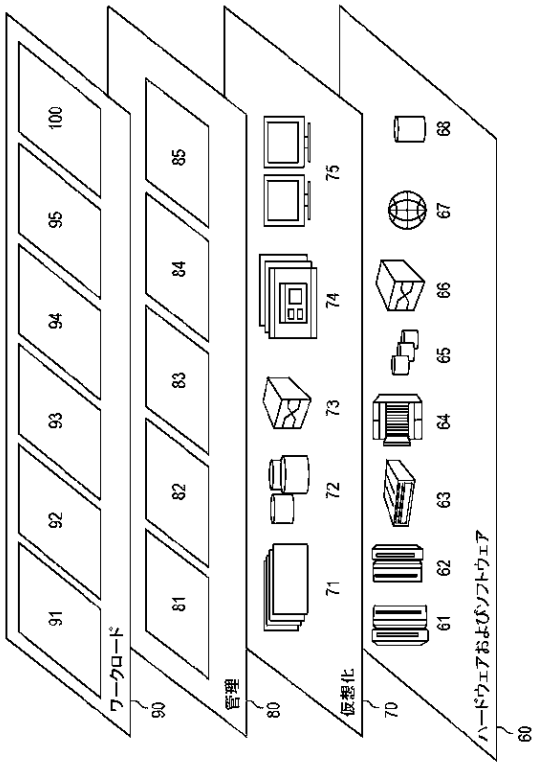
【図９】



【図１０】



【図１１】


ワークロード
90
91
92
93
94
95
100
管理
80
81
82
83
84
85
仮想化
70
71
72
73
74
75
ハードウェアおよびソフトウェア
60
61
62
63
64
65
66
67
68

フロントページの続き

(72)発明者　バイド、マニシュ、アナンド
インド５０００８１　テランガーナ州　ハイデラバード　マダプール　ハイテクシティ　マインドスペース
(72)発明者　ロヒア、プラネイ、クマール
インド５６００４５　カルナータカ州　バンガロール　ラケナハリ・アンド・ナガワラ・ビレッジ　アウター・リング・ロード　マニヤタ・エンバシー・ビー
(72)発明者　ナテサン　ラママーシー、カーティケヤン
アメリカ合衆国１０５９８　ニューヨーク州　ヨークタウン・ハイツ　ピーオー・ボックス２１８　キッチャワン・ロード　１１０１
(72)発明者　プリ、ルチル
アメリカ合衆国１０５９８　ニューヨーク州　ヨークタウン・ハイツ　ピーオー・ボックス２１８　キッチャワン・ロード　１１０１
(72)発明者　サハ、ディプティカルヤン
インド５６００４５　カルナータカ州　バンガロール　ラケナハリ・アンド・ナガワラ・ビレッジ　アウター・リング・ロード　マニヤタ・エンバシー・ビー
(72)発明者　ヴァーシュニー、クシュ、ラージ
アメリカ合衆国１０５９８　ニューヨーク州　ヨークタウン・ハイツ　ピーオー・ボックス２１８　キッチャワン・ロード　１１０１

審査官　武田　広太郎

(56)参考文献　特開２０１４－２１１７６１（ＪＰ，Ａ）
KAMIRAN, Faisal, Decision Theory for A 式外 Discrimination-aware Classification, Proceedinds of 12th International Conf. on Data Mining, 2013年01月17日, web.lums.edu.pk/ akarim/pub/decision_theory_icdm2012.pdf

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｎ　２０／００
Ｇ０６Ｑ　１０／０４