(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7043512号
(P7043512)

(45)発行日　令和4年3月29日(2022.3.29)

(24)登録日　令和4年3月18日(2022.3.18)

(51)国際特許分類

Ｇ０６Ｑ　10/04　(2012.01)		Ｇ０６Ｑ	10/04	
Ｇ０６Ｎ　20/00　(2019.01)		Ｇ０６Ｎ	20/00	１３０

ＦＩ

請求項の数　6　（全40頁）

(21)出願番号　特願2019-554997(P2019-554997)
(86)(22)出願日　平成30年4月2日(2018.4.2)
(65)公表番号　特表2020-517004(P2020-517004A)
(43)公表日　令和2年6月11日(2020.6.11)
(86)国際出願番号　PCT/US2018/025661
(87)国際公開番号　WO2018/191048
(87)国際公開日　平成30年10月18日(2018.10.18)
審査請求日　令和2年11月19日(2020.11.19)
(31)優先権主張番号　62/485,314
(32)優先日　平成29年4月13日(2017.4.13)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/820,316
(32)優先日　平成29年11月21日(2017.11.21)
最終頁に続く

(73)特許権者　502303739
オラクル・インターナショナル・コーポレイション
アメリカ合衆国カリフォルニア州９４０６５レッドウッド・シティー，オラクル・パークウェイ５００
(74)代理人　110001195
特許業務法人深見特許事務所
(72)発明者　アッバス，フセイン
アメリカ合衆国、９４１０２ カリフォルニア州、サン・フランシスコ、テイラー・ストリート、５５５、アパートメント・５０２
審査官　岡 裕之
最終頁に続く

(54)【発明の名称】　パイプ漏れを予測する新規な自律的人工知能システム

(57)【特許請求の範囲】
【請求項１】
パイプ漏れを予測するための、コンピュータによって実施される方法であって、
第１のデータ項目および第１の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む訓練用データセットにアクセスすることを備え、前記第１のデータ項目は前記第１の複数のパイプのそれぞれのパイプの特性を含み、前記パイプの特性は、前記パイプの長さと、前記パイプの土壌抵抗率と、前記パイプの最大圧力定格と、前記パイプの標高とを含み、前記方法はさらに、
教師付き機械学習技術を適用して、パイプの漏れ予測を判断するように構成される予測モデルを、前記第１の複数のパイプのそれぞれのパイプに関連付けられる前記第１のデータ項目に基づいて訓練することにより生成することを備え、前記予測モデルは、ランダムフォレストモデルを含み、前記方法はさらに、
第２のデータ項目および第２の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む検証データセットにアクセスすることを備え、前記第２のデータ項目は前記第２の複数のパイプのそれぞれの前記パイプの特性を含み、前記方法はさらに、
前記予測モデルを検証することを備え、前記予測モデルを検証することは、少なくとも、前記予測モデルを前記第２のデータ項目に適用することにより、前記第２の複数のパイプのそれぞれのパイプの漏れ予測のセットを判断することと、
前記第２の複数のパイプのそれぞれのパイプの漏れ予測を前記第２の複数のパイプのそれぞれのパイプの既知の漏れと比較して、前記第２の複数のパイプのそれぞれのパイプの漏

れ予測の正答率を判断することと、
前記第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に基づいて混同行列を生成することと、
前記混同行列に基づいて、前記第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に関連付けられる真陽性率、偽陽性率、真陰性率、および偽陰性率を判断することと、
前記真陽性率が第１のしきい値を超えていると判断することと、
前記真陰性率が第２のしきい値を超えていると判断することとによって行われ、前記方法はさらに、
第３の複数のパイプに関連付けられる第３のデータ項目を含むパイプラインデータセットにアクセスすることを備え、前記第３のデータ項目は前記第３の複数のパイプのそれぞれの前記パイプの特性を含み、前記方法はさらに、
前記予測モデルを前記パイプラインデータセットに適用して、前記第３の複数のパイプのそれぞれのパイプの漏れ予測を判断することを備える、方法。
【請求項２】
前記第１のしきい値は、前記予測モデルの適用からの前記第２の複数のパイプのそれぞれのパイプの漏れ予測のセットの統計分布に基づいて自動的に判断される、請求項１に記載の方法。
【請求項３】
前記第２のしきい値は、前記予測モデルの適用からの前記第２の複数のパイプのそれぞれのパイプの漏れ予測のセットの統計分布に基づいて自動的に判断される、請求項１または２に記載の方法。
【請求項４】
前記第３の複数のパイプのそれぞれのパイプの判断された漏れ予測に基づいて、前記第３の複数のパイプのそれぞれのパイプをオーダーすることをさらに備える、請求項１～３のいずれか１項に記載の方法。
【請求項５】
コンピューティングシステムであって、
１つ以上のデータストアを備え、前記１つ以上のデータストアは、
第１のデータ項目および第１の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む訓練用データセットを格納し、前記第１のデータ項目は前記第１の複数のパイプのそれぞれのパイプの特性を含み、前記パイプの特性は、前記パイプの長さと、前記パイプの土壌抵抗率と、前記パイプの最大圧力定格と、前記パイプの標高とを含み、前記１つ以上のデータストアはさらに、
第２のデータ項目および第２の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む検証データセットを格納し、前記第２のデータ項目は前記第２の複数のパイプのそれぞれの前記パイプの特性を含み、前記コンピューティングシステムはさらに、
コンピュータプロセッサと、
コンピュータ読取可能記憶媒体とを備え、前記コンピュータ読取可能記憶媒体は命令を格納し、前記命令は、前記コンピュータプロセッサによって実行されると、前記コンピュータプロセッサに、
前記訓練用データセットにアクセスさせ、
教師付き機械学習技術を適用させて、パイプの漏れ予測を判断するように構成される予測モデルを、前記第１の複数のパイプのそれぞれのパイプに関連付けられる前記第１のデータ項目に基づいて訓練することにより生成させ、
前記検証データセットにアクセスさせ、
前記予測モデルを検証させるよう構成され、前記予測モデルは、ランダムフォレストモデルを含み、前記予測モデルを検証することは、少なくとも、
前記予測モデルを前記第２のデータ項目に適用することにより、前記第２の複数のパイプのそれぞれのパイプの漏れ予測のセットを判断することと、
前記第２の複数のパイプのそれぞれのパイプの漏れ予測を前記第２の複数のパイプのそれ

ぞれのパイプの既知の漏れと比較して、前記第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率を判断することと、
前記第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に基づいて混同行列を生成することと、
前記混同行列に基づいて、前記第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に関連付けられる真陽性率、偽陽性率、真陰性率、および偽陰性率を判断することと、
前記真陽性率が第１のしきい値を超えていると判断することと、
前記真陰性率が第２のしきい値を超えていると判断することとによって行われ、前記命令は、さらに、前記コンピュータプロセッサによって実行されると、前記コンピュータプロセッサに、
第３の複数のパイプに関連付けられる第３のデータ項目を含むパイプラインデータセットにアクセスさせ、
前記予測モデルを前記パイプラインデータセットに適用させて、前記第３の複数のパイプのそれぞれのパイプの漏れ予測を判断させるよう構成され、
前記第３のデータ項目は前記第３の複数のパイプのそれぞれの前記パイプの特性を含む、
コンピューティングシステム。
【請求項６】
コンピュータプロセッサに請求項１～４のいずれか１項に記載の方法を実行させるためのコンピュータ読取可能プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願への相互参照
この出願は、「A NOVEL AUTONOMOUS ARTIFICIALLY INTELLIGENT SYSTEM TO PREDICT PIPE LEAKS（パイプ漏れを予測する新規な自律的人工知能システム）」と題される２０１７年１１月２１日に出願された米国非仮出願第１５／８２０，３１６号および２０１７年４月１３日に出願された米国仮出願第６２／４８５，３１４号の利益ならびに優先権を主張し、その全体をすべての目的のためにここに引用により援用する。
【背景技術】
【０００２】
背景
公益企業は、多くの場合、パイプを使用して公益を伝達および配送する（例：水を配送するパイプ）。しかしながら、時間の経過とともに、これらのパイプはさまざまな理由で故障し、漏れが発生し始める可能性がある。これらの漏れは、多くの理由で公益企業にとって非常にコストがかかる。公益企業は、漏れの損失（例えば、パイプから漏れる水のコスト）または漏れに関連する損害（例えば、水に起因する洪水）を負担しなければならないかもしれない。また、一部のパイプは簡単にアクセスできず（たとえば、地下に埋められている）、場合によっては漏れがすぐには分からないかもしれないため、漏れの識別および修理に関連するコストもかかる。パイプの漏れは、水道本管の破損や陥没穴を引き起こし、それは、環境や公衆衛生への危険に加えて、周囲の事業への経済的損害、税金の浪費、交通遅延、機会費用を引き起こす。漏れているパイプを修理するには、公益企業は水を他のパイプに経路付けし直す必要がある。これにより、これらのパイプに大量の追加のストレスがかかり、非常に広い地理的領域および元のパイプ漏れから遠く離れた場所でパイプ漏れの連鎖反応を引き起こす可能性がある。さらに、特定のパイプが漏れるかどうかを予測する機能により、公益企業は、漏れの可能性を減らす将来のパイプの建設入札契約の新たな製造業者購入基準を設計および実施できる。
【０００３】
これらの公益企業は、どのパイプが漏れるのかを事前に予測できることから多大な利益を得、なぜならば、それにより、企業はそれらの資源を、これらの漏れが発生する前に止めることに充てることができるであろうからである。しかしながら、実際には、これを行う

のは非常に困難であり得る。さまざまな例において、パイプはさまざまな寸法および材質のものであり、さまざまな場所および方法で設置されている。言い換えれば、特定のパイプが漏れるかどうかは、多くの因子によって決定され得る。

【０００４】

したがって、パイプ漏れに影響を与える多数の因子を考慮に入れることができる、信頼性が高く正確なパイプ漏れ予測システムが必要とされている。本開示の実施形態は、少なくともこれらのニーズに対処することに向けられている。

【発明の概要】

【課題を解決するための手段】

【０００５】

簡単な要約

本開示に記載される実施形態は、パイプ（例えば、水などの何らかの物質を運ぶ公益事業設備パイプ）が漏れそうであるかどうかを予測するように構成されたパイプ漏れ予測システムに向けられる。パイプ漏れ予測システムは、１つ以上の機械学習技術に基づいた予測モデルを含み得る。たとえば、予測モデルは、出力をランダムフォレストモデルなどのカテゴリ（たとえば、漏れありまたは漏れなし）に分類するための教師付き機械学習法を用いて生成できる。予測モデルは、訓練用データセットに含まれるパイプについてのデータを用いて訓練でき、訓練用データセットは、さまざまなパイプの特性に関する情報（例えば、それらのパイプの寸法、それらのパイプの材料、それらのパイプの年齢、それらのパイプの位置など）、およびそれらのパイプが漏れているかどうかについての知識を含むことができる。予測モデルは、訓練用データセットのデータを用いて、漏れのないパイプに関連付けられるパターンおよび漏れのあるパイプに関連付けられるパターンを判断することができる。これらのパターンを任意のパイプの特性に適用して、そのパイプが漏れるかどうかを予測できる。

【０００６】

いくつかの実施形態においては、予測モデルが生成されると、パイプ漏れ予測をそれらのパイプのうちどれが実際に漏れたかの知識と比較することにより、予測モデルを検証データセットに対して検証することができる。その後、これらの結果に基づいて混同行列を生成し、それを解釈して、予測モデルに関する貴重な洞察を得ることができる。予測モデルが検証されると、予測モデルを任意のパイプに適用して、そのパイプが漏れるかどうかを予測し得る。漏れそうであると識別されたパイプは、潜在的な修理または予防保守のためにさらなる調査に割り当てることができる。

【０００７】

いくつかの実施形態では、パイプ漏れを予測するためのコンピューティングシステムおよび／またはコンピュータによって実施される方法が開示される。たとえば、この方法は、第１のデータ項目および第１の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む訓練用データセットにアクセスすることを備え得、第１のデータ項目はそれぞれのパイプの特性を含む。教師付き機械学習技術をこの訓練用データセットに適用して予測モデルを生成し、この予測モデルは、第１の複数のパイプのそれぞれのパイプに関連付けられる第１のデータ項目に基づいて訓練されることによりパイプの漏れ予測を判断するように構成される。予測モデルが生成されると、この方法は、さらに、第２のデータ項目および第２の複数のパイプのそれぞれのパイプに関連付けられる既知の漏れを含む検証データセットにアクセスすることを備え、第２のデータ項目はそれぞれのパイプの特性を含む。検証データセットを用いて、予測モデルを第２のデータ項目に適用することにより第２の複数のパイプのそれぞれのパイプの漏れ予測のセットを判断することによって予測モデルを検証することができる。第２の複数のパイプのそれぞれのパイプの漏れ予測を第２の複数のパイプのそれぞれのパイプの既知の漏れと比較して、第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率を判断することができる。その後、第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に基づいて混同行列を生成することができる。混同行列から、第２の複数のパイプのそれぞれのパイプの漏れ予測の正答率に関連付けられ

る真陽性率および真陰性率を判断することができる。予測モデルが有効であると判断される（たとえば、真陽性率が第１のしきい値を超えると判断され、真陰性率が第２のしきい値を超えると判断される）と、この方法は、さらに、第３の複数のパイプに関連付けられる第３のデータ項目を含むパイプラインデータセットにアクセスすることを備える。検証された予測モデルをこのパイプラインデータセットに適用して、第３の複数のパイプのそれぞれのパイプの漏れ予測を判断することができる。

【図面の簡単な説明】

【０００８】

【図１】本開示の実施形態によるパイプ漏れ予測のシステム図を示す。

【図２】本開示の実施形態によるパイプ漏れ予測のフローチャートを示す。

【図３】本開示の実施形態によるさまざまなパイプに関連付けられた例示的なデータを示す。

【図４】本開示の実施形態によるパイプデータベース内のパイプレコードの例示的な数字を示す。

【図５】本開示の実施形態によるランダムフォレストモデルの混同行列を示す。

【図６】本開示の実施形態によるランダムフォレストモデルの変数重要度プロットを示す。

【図７】本開示の実施形態によるロジスティック回帰の確率の統計的分布を示す。

【図８】本開示の実施形態によるロジスティック回帰の混同行列を示す。

【図９】本開示の実施形態によるロジスティック回帰の係数のリストを示す。

【図１０】本開示の実施形態による単純ベイズモデルの混同行列を示す。

【図１１】本開示の実施形態によるパイプ漏れ予測を実施するためのブロック図を示す。

【図１２】本開示の実施形態によるパイプ漏れ予測のためのハイブリッドシステム図を示す。

【図１３】実施形態の１つを実施するための分散型システムの簡略図を示す。

【図１４】本開示の実施形態に従って、実施形態システムのコンポーネントによって提供されるサービスをクラウドサービスとして提供することができるシステム環境のコンポーネントの簡略ブロック図である。

【図１５】本発明のさまざまな実施形態を実施することができる例示的なコンピュータシステムを示す。

【発明を実施するための形態】

【０００９】

詳細な記載

以下の記載では、説明の目的で、本発明の実施形態の完全な理解のために、特定の詳細が述べられる。しかしながら、これらの特定の詳細なしでさまざまな実施形態を実施できることは明らかであろう。図および記載は、限定することを意図したものではない。

【００１０】

いくつかの図に描かれているシステムは、さまざまな構成で提供され得る。特定の実施形態では、システムは、システムの１つ以上のコンポーネントがクラウドコンピューティングシステムの１つ以上のネットワークにわたって分散される分散型システムとして構成されてもよい。特定の実施形態では、システムは、仮想環境または非仮想環境で動作するように構成されてもよい。

【００１１】

導入

パイプ漏れ予測システムは、パイプが故障する可能性があるかどうかを事前に予測するために、公共企業（例：水、ガス、下水などを輸送する企業）またはエネルギー探査／生産企業（例：パイプを介して石油、ガス、燃料を輸送する企業）など、パイプを用いる任意の実体によって用いられ得る。この予測は、パイプのメンテナンスを実行するなどの予防策を講じるために使用できる。

【００１２】

パイプ漏れを予測するいくつかの従来の方法は理論に基づいており、経験的データではな

く仮定に依存し得る。例えば、パイプ漏れ予測システムの一実施形態は、過去のパイプ漏れの履歴などの因子が最も重要な因子であると想定している。言い換えれば、パイプが過去にすでに漏れているかどうかは、パイプが将来再び漏れるかどうかの最も重要な決定因子である。しかしながら、このようなシステムは、他の重要な因子を見落とす可能性があり、経験的に検証または完全に理解されていない仮定に依存する可能性もある。たとえば、あるシステムは、以前に漏れたパイプは再び漏れる可能性が高いと正しく仮定するかもしれないが、そのようなシステムは、そもそもそのパイプが漏れた理由についての説明を提供しない。したがって、そのようなシステムは、その予測が、パイプが以前に漏れたことに非常に依存しているため、パイプにおける最初の漏れの発生を予測するのにはあまり役に立たない。

【００１３】

代わりに、本開示のさまざまな実施形態は、パイプ漏れを予測するように設計された、データに基づくパイプ漏れ予測システムに向けられる。このようなシステムは、パイプ漏れに関する任意の利用可能なデータをして語らしめ、このシステムは、重要な因子または重要でない因子を自身で判断できる。いくつかの実施形態では、このパイプ漏れ予測システムは、異なる因子の相対的な重要度およびそれらの因子間の任意の関係（たとえば、相関／原因）を判断することもできる場合がある。この判断を用いて、パイプの初回漏れを正確に予測し、パイプ漏れの根本的な原因をより深く理解することができる。

【００１４】

いくつかの実施形態では、パイプ漏れ予測システムは、大量のデータを利用するように設計されてもよい。いくつかの実施形態では、パイプ漏れ予測システムは、パイプ（たとえば、多数のパイプのデータを考慮する）および因子（たとえば、パイプ漏れに影響を与える可能性のある多種多様な因子のデータを考慮する）に関してできるだけ多くのデータを考慮し得る。したがって、システムは数千または数百万のパイプのデータを考慮し、各パイプの多数の属性または因子に関するデータがある場合がある。より多くのパイプおよび因子が追跡されるにつれて、システムによって考慮されるデータのサイズは大幅に増加する。場合によっては、データの合計サイズがギガバイトまたはテラバイトを超えることがあり、人間がすべてのデータを利用して暗算やペンと紙の計算を行うことは不可能であろう。本開示の実施形態は、他の方法を免れるであろうパイプ漏れの多数のケースを正確に予測するために、膨大な量のデータのデータ駆動型分析を可能にする。

【００１５】

いくつかの実施形態では、パイプ漏れ予測システムは、さまざまな機械学習アルゴリズムを利用してもよい。特に、パイプ漏れ予測システムは、既知のパイプ漏れに関する既存のデータを用いて予測モデルを訓練するように、教師付き機械学習技術を用い得る。このような教師付き機械学習技術には、限定ではなく、例として、分析学習、人工ニューラルネットワーク、逆伝播、ブースティング（メタアルゴリズム）、ベイジアン統計、事例ベース推論、決定木学習、帰納論理プログラミング、ガウス過程回帰、データ処理のグループ法、カーネル推定器、学習オートマトン、学習分類器システム、最小メッセージ長（決定木、決定グラフなど）、多重線形部分空間学習、単純ベイズ分類器、最大エントロピー分類器、条件付きランダムフィールド、最近傍アルゴリズム、確率的で近似的に正しい学習（ＰＡＣ）学習、リップルダウンルール、サポートベクターマシン、最小複雑度マシン（ＭＣＭ）、ランダムフォレスト、分類器のアンサンブル、順序分類、データの前処理、および統計的関係学習が含まれる。

【００１６】

いくつかの実施形態では、パイプ漏れ予測システムは、パイプについての入力データを用いて各パイプを１つ以上のカテゴリ（たとえば、漏れそうである、または漏れそうでない）に分類する分類ベースの教師付き機械学習技術を利用し得る。分類タイプの教師付き機械学習技術の例には、サポートベクターマシン（ＳＶＭ）、ニューラルネットワーク、単純ベイズ分類器、決定木、適応的ブースティング（Ａｄａｂｏｏｓｔ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、判別分析、最近傍法（ｋＮＮ）が含まれる。

【００１７】
いくつかの実施形態では、パイプ漏れ予測システムは、パイプについての入力データを用いてさまざまなパイプの連続応答値（たとえば、パイプが漏れそうになる前の漏れの確率または時間の量）を計算するために、回帰ベースの教師付き機械学習技術を利用することがある。回帰タイプの教師付き機械学習技術の例には、線形回帰、投げ縄回帰、リッジ回帰、エラスティックネット回帰、部分最小二乗回帰、多項式回帰、ランダムフォレスト、ＳＶＭ、ＸＧＢｏｏｓｔ、Ａｄａｂｏｏｓｔ、非線形回帰、一般化線形モデル、決定木、ニューラルネットワークなどが含まれる。いくつかの実施形態では、パイプ漏れ予測システムは、時間を考慮した予測を行うように構成されてもよい。言い換えれば、システムは、パイプがいつ故障するかだけでなく、故障する時間を予測し得る。たとえば、システムは、パイプが１００日以内に故障すると予測し得る。
【００１８】
いくつかの実施形態では、パイプ漏れ予測システムは、異なる予測モデルまたは機械学習技術の組み合わせを利用してもよい。さまざまな機械学習技術にはさまざまな利点および欠点があるため、異なる目的に利用され得る。たとえば、ある機械学習技術を用いて、ある結果または情報のセットを取得でき、別の機械学習技術を用いて、他方の技術では取得できない別の結果または情報のセットを取得できる。場合によっては、ある機械学習技術を用いて、別の機械学習技術の結果を検証できる。場合によっては、複数の機械学習技術をアンサンブルで組み合わせて、コンポーネント手法を超えて予測正答率を高めることができる。
【００１９】
本開示では、機械学習技術の組み合わせとして、ランダムフォレスト、一般化線形モデル（例えば、ロジスティック回帰モデル）、および単純ベイズを用いるパイプ漏れ予測システムの実施形態が開示される。しかしながら、この組み合わせは例示のみを目的としており、本願においては、理解を容易にする目的で、継続的に参照される。機械学習アルゴリズムの他の好適な組み合わせが使用されてもよい。
【００２０】
例示的な実施の形態
図１は、本開示の実施形態によるパイプ漏れ予測のシステム図を示す。
【００２１】
いくつかの実施形態では、パイプ設置業者１０２は、公益企業などの実体に代わってパイプ（図示せず）を設置する業務を課され得る。パイプ設置業者１０２は、パイプを設置し、次いで、設置されたパイプに関する詳細をコンピューティングデバイス１０４を介してパイプ統合システム１０６に提供することができる。例えば、パイプ設置業者１０２は、山腹にパイプを設置し、次いでコンピューティングデバイス１０４を用いて、パイプの位置、パイプの寸法および材料、パイプの目的およびそれが輸送する内容物などに関する情報など、パイプに関連する詳細または属性をパイプ統合システム１０６に提供することができる。この情報は、特定のパイプが漏れるかどうかを判断する際にパイプ漏れ予測システムによって考慮され得る多数の因子についてのデータを含み得、パイプの詳細／属性の他の例を図３に示す。パイプ統合システム１０６は、パイプに関するこのデータをパイプデータベース１０８に格納することができ、パイプデータベース１０８は、公益企業によって設置されたさまざまなパイプ（例えば、パイプ１２０－１から１２０－Ｎ）のデータを含むことができる。
【００２２】
いくつかの実施形態では、パイプの特定の詳細がすでに既知である場合があるため、パイプ設置業者１０２はこの情報をパイプ統合システム１０６に供給する必要がない場合がある。例えば、パイプ設置業者１０２は、パイプ製造業者に設置すべきパイプをオーダーする業務を課されることもあり得る。したがって、パイプに関する特定の詳細をオーダーから取得して、パイプ統合システム１０６に提供することができ、そのパイプに関する情報がパイプデータベース１０８ですでに利用可能になる。しかし、パイプ設置業者１０２は

、コンピューティングデバイス１０４を用いてパイプに関連付けられる情報の一部を更新しなければならない場合がある（例えば、パイプの設置中に、設置場所の条件が予想とは異なることになった、など）。各パイプに関連付けられた情報は、状況が変化するにつれて時間とともに更新する必要がある場合もある。例えば、パイプデータベース１０８は、各パイプが漏れたかどうかに関する情報を含み得る。漏れのないパイプが漏れを発生させた場合、その情報はパイプデータベース１０８内で更新する必要があるであろう。この更新は、コンピューティングデバイス１０４を用いるパイプ設置業者１０２などによって、パイプ統合システム１０６を介して、または現場で、行なわれ得る。

【００２３】

したがって、パイプデータベース１０８は、パイプ設置業者１０２によって設置されたパイプを含む、公益企業によって設置されたさまざまなパイプの、時間とともに変化し更新される、膨大な量のデータを含み得る。これは、図において、パイプ１２０－１から１２０－Ｎに関する情報を含むように示されるパイプデータベース１０８のコンテンツによって表されている。この情報には、各パイプの仕様、各パイプの目的およびそれによって輸送される内容物などが含まれ得る。パイプデータベース１０８のデータは、パイプ漏れ予測システム１１０のための訓練用セットとして用いられ得、それは、漏れを発生させたすべてのパイプ間のパターンを判断するために、さまざまなパイプに関連付けられる情報のすべてを使用し得る。これらのパターンに加えて、パイプ漏れ予測システム１１０は、パイプが漏れを発生させるかどうかを決定するさまざまな因子間の関係を識別し、その情報を用いて予測モデルを開発することもできる。

【００２４】

いくつかの実施形態では、パイプ漏れ予測システム１１０は、パイプの故障または漏れの発生を予測するために用いることができる。これは、（例えば、パイプ１２０－１～１２０－Ｎの関連情報が時間とともに更新されるにつれ）パイプデータベース１０８内のパイプ１２０－１～１２０－Ｎに関する予測を行うことを含み得る。パイプ漏れ予測システム１１０は、パイプデータベース１０８にまだ導入されていない新たに設置されたパイプなど、パイプデータベース１０８の外部のパイプについて予測することもできる。例えば、パイプ設置業者１０２がパイプを設置し（または設置直前でも）、パイプに関連付けられる入力情報をパイプ漏れ予測システム１１０に提供した後、システムは、そのパイプが故障しそうかどうか、そのパイプがいつ故障しそうか、パイプが故障しそうな確率など、を予測または推定し得る。いくつかの実施形態では、パイプ漏れ予測システム１１０は、コンピューティングデバイス１１２を介して、調査員１１４に、漏れそうであると識別されたパイプのリストを提供することができる場合がある。調査員１１４は、その情報を用いて、漏れそうであると判断されたパイプを調べて、予防保守を実行する必要があるかどうかを確認でき、または場合によってはすでに漏れ始めているパイプを置換する。

【００２５】

図２は、本開示の実施形態によるパイプ漏れ予測のフローチャートを示す。

ブロック２０２で、訓練用データセット２２０内に含まれるパイプについてのデータを用いて予測モデルを生成することができる。訓練用データセット２２０は、パイプデータベース１０８などの、さまざまなパイプについてのデータを含む、より大きなデータセットのサブセットであってもよい。より大きなデータセットは、予測モデルを訓練するために用いられる訓練用データセット２２０、および訓練された予測モデルの結果を検証するために用いられる検証データセットに、ランダムに分割され得る。訓練用データセット２２０は、さまざまなパイプの特性に関する情報（例えば、それらのパイプの寸法、それらのパイプの材料、それらのパイプの年齢、それらのパイプの位置など）を含むことができる。さらに、訓練用データセット２２０は、それらのパイプのどれが漏れを有していると知られているかに関する情報も含み得る。したがって、予測モデルは、漏れのないパイプに関連付けられたパターンおよび漏れのあるパイプに関連付けられたパターンを判断するために、訓練用データセット２２０のデータを用いることができる。モデルは、これらのパイプの特性が提供されている場合、これらのパターンを学習および使用して、パイプの漏

れを予測し得る。予測モデルは、前述のようなさまざまな機械学習アルゴリズムを用いて生成でき；たとえば、予測モデルは、ランダムフォレストモデル、ロジスティック回帰モデル、単純ベイズモデルなどであり得る。

【００２６】

ブロック２０４で、予測モデルが生成されると、モデルを検証することができる。予測モデルの検証には、サブブロック２０６、２０８、および２１０などの１つ以上のサブブロックを含めることができる。ブロック２０６で、予測モデルを用いて、検証データセットからのパイプデータを用いてパイプ漏れを予測することができる。たとえば、予測モデルは入力として検証データセット内のさまざまなパイプの特性を取得し、それらを用いてそれらのパイプのいずれが漏れそうかを識別する。同時に、検証データセットには、これらのパイプのどれが実際に漏れたかに関するデータが含まれることになる。したがって、漏れがあると予測されたパイプのセットを実際に漏れのあるパイプのセットと比較して、モデルの正答率を判断することができる。ブロック２０８において、パイプの既知のステータス（例えば、漏れありまたは漏れなし）に対する予測モデルの結果の比較に基づいて混同行列を生成することができる。この混同行列は、予測モデルによって識別された真陽性、真陰性、偽陽性、および偽陰性の量を示すことができ、それをさらに用いて、予測モデルに関連付けられる正答率を計算することができる。ブロック２１０では、混同行列を解釈して、予測モデルに関する貴重な洞察を得ることができる。たとえば、混同行列によって、モデルが高い正答率を有していると示される場合、モデルに関連付けられる基礎となる仮定が正しい可能性がある。いくつかの実施形態では、機械学習モデルから人間が読み取れる形式への推論の自動翻訳があり得る（例えば、レポートを、現場の調査員に送信できるＰＤＦ形式またはテキストメッセージで生成できる）。これにより、人間は、機械学習アルゴリズムから生じる洞察に集中することができ、かなりの統計／数学の専門知識が必要になる場合があるそれらの数値出力を理解する必要がない。いくつかの実施形態では、ユーザは、モデルの詳細に飛び込んで自分で推論を引き出すか、またはパイプ漏れ予測システムによって直接提供されるモデルからの詳細／推論を得るオプションを有してもよい。

【００２７】

ブロック２１２で、予測モデルが検証されると、予測モデルは、次いで、新たなコホート（例えば、訓練用データセットまたは検証データセットにないパイプ）に適用され得る。例えば、予測モデルは、さまざまなパイプの特性についてのデータを含むパイプラインデータセット２２２のパイプデータに適用できる。予測モデルは、パイプラインデータセット２２２内のどのパイプが漏れそうかを識別するために用いられるであろう。

【００２８】

ブロック２１４で、パイプラインデータセット２２２に適用された予測モデルの結果を解釈することができる。例えば、予測モデルは、パイプラインデータセット２２２内の漏れそうなパイプのサブセットを特定したかもしれない。これらのパイプは、潜在的な修理または予防保守のためにさらに調査され得る。いくつかの実施形態では、漏れそうなこれらのパイプの識別は、調査員１１４など、パイプの調査および予測結果の検証を任された調査員に与えられ得る。パイプ漏れ予測システムは、コンピューティングデバイス１１２を介してパイプラインデータセット２２２内の漏れそうなパイプを調査員１１４に通知し得、調査員１１４は彼／彼女のリソースを用いてそれらの識別されたパイプを調査することができる。

【００２９】

図３は、本開示の実施形態によるさまざまなパイプに関連付けられた例示的なデータを示す。

【００３０】

より具体的には、図３は、列３０２に示される「fid」（例えば、パイプＩＤ）および列３０４に示される「wsm_nbr」（例えば、追跡番号）に基づいて識別される６つのパイプに関連する生データを示す。たとえば、第１のパイプのfidは「9277583」で、wsm_nbr

は「172-147」である。パイプごとにさまざまな種類の情報を利用できる。しかしながら、このリストは網羅的であることを意味するものではなく、図１のパイプデータベース１０８にデータをどのように格納できるかを示すなどの例示目的のためのものであり；パイプに関連付けられる他のデータを含めて、予測モデルの生成時にパイプ漏れ予測システムで考慮することができる。

【００３１】

いくつかの実施形態において、データは、パイプがどのように使用され得るかを示すカテゴリ変数を示す列３０６（ヘッダ「use_type（使用タイプ）」で示される）を含み得る。パイプについてのさまざまな使用タイプについての変数の例には、ＤＩＳＴ（例：配給）、ＴＲＵＮＫ（例、供給幹線）、ＩＲＲＩＧ（例、灌漑）、ＤＥＣＯＭ（例、分解管）、ＷＥＬＬ（例、井戸）、ＣＨＬＯＲ（例、塩素）、ＤＲＡＩＮ（例、排水管）、ＡＭＭＯＮＩＡ（例、アンモニア）、ＲＥＣＬ（例、再生水管）、ＭＡＮＩＦＯＬＤ（例、マニホールド管）、ＬＳＬＡＴＥＲＡＬ（例、個人所有の下水道）、および＃ＮＵＬＬ＃（パイプの使用が不明であるか、または使用タイプに関する他のカテゴリの１つに準拠していないことを示す場合がある）が含まれる。たとえば、列３０６は、６つのパイプすべてについてＤＩＳＴを示し、６つのパイプの各々が配給に用いられることを示す。さまざまなパイプの使用タイプを追加することで、パイプ漏れ予測システムは、特定の使用タイプがパイプで漏れを引き起こす可能性がより高いかどうかを判断でき得る。

【００３２】

いくつかの実施形態では、データは、パイプの幅を示す数値変数を示す列３０８（ヘッダ「pipe_diam（パイプ直径）」で示される）を含むことができる。任意の測定単位を事前に指定して用いることができる。図において、列３０８の数字は、各パイプの幅をインチで表してもよい。例えば、列３０８は、第１のパイプに関連付けられた値「１２」を有し、これは、第１のパイプが１２インチの直径を有することを示し得る。各パイプの直径を知ることにより、パイプ漏れ予測システムは、パイプの直径がパイプ漏れの可能性をどのように増加または減少させ得るかを判断し得る。

【００３３】

いくつかの実施形態では、データは、パイプがどれくらいの長さであるかを示す数値変数を意味する列３１０（ヘッダ「len_ft」で示される）を含むことができる。任意の測定単位を事前に指定して用いることができる。図において、列３１０の数字は、各パイプの長さをフィートで表してもよい。たとえば、列３１０の第１のパイプに関連付けられる値は３２９．７６であり、これは、第１のパイプの長さがほぼ３３０フィートであることを示し得る。各パイプの長さを知ることにより、パイプ漏れ予測システムは、パイプの長さがパイプ漏れの可能性をどのように増加または減少させ得るかを判断し得る。

【００３４】

いくつかの実施形態では、データは、パイプの壁がどのくらいの厚みであるかを示す数値変数を示す列３１２（ヘッダ「thickness（厚み）」で示される）を含んでもよい。任意の測定単位を事前に指定して用いることができる。図において、列３１２の数字は、各パイプの壁の厚みをインチで表してもよい。例えば、列３１２は、第１のパイプに関連付けられる０．７１の値を有し、これは、第１のパイプが４分の３インチの厚みを有することを示し得る。各パイプの壁の厚みを知ることにより、パイプ漏れ予測システムは、パイプの壁厚がパイプ漏れの可能性をどのように増加または減少させ得るかを判断し得る。常識では、より厚い壁はパイプ漏れの可能性を減じることを示唆するが、常識では、他の因子と比較して壁の厚みの相対的な重要度を識別することはできない。たとえば、パイプの直径が半分になり、壁の厚みが３分の１に減じられた場合、常識では、漏れの可能性が増加したか減少したかを知らせるのには役立たないであろう。

【００３５】

いくつかの実施形態では、データは、パイプの材料を示すカテゴリ変数を示す列３１４（ヘッダ「material（材料）」で示される）を含むことができる。パイプのさまざまな材料の変数の例には、ＣＩ（例、鋳鉄）、ＳＴＬ（例、鋼）、ＡＣ（例、アスベストセメント

）、ＤＩ（例、ダクタイル鋳鉄）、ＣＯＰ（例、銅）、ＨＤＰＥ（例、高密度ポリエチレン）、ＣＯＮＣ（例、コンクリート）、ＰＶＣ（例、ポリ塩化ビニル）、ＣＭ（例、波形金属）、ＢＲＳ（例、真鍮）、ＰＰ（例、ポリプロピレン）が含まれる。例えば、列３１４は、第１のパイプに対してＣＩを示し、これは第１のパイプが鋳鉄で作られていることを示すことができる。さまざまなパイプについてさまざまな材料を追加することで、パイプ漏れ予測システムは、特定の材料の使用がパイプの漏れを引き起こす可能性がより高いかどうかを判断し得る。

【００３６】

いくつかの実施形態では、データは、パイプのライニングが取り替えられているかどうかを示すカテゴリ変数を示す列３１６（ヘッダ「relined（ライニングが取り替えられた）」で示される）を含むことができる。変数の例には、ＣＬおよび＃ＮＵＬＬ＃（パイプのライニングが取り替えられていないことを示し得る）が含まれる。例えば、列３１６は、第１のパイプに対して＃ＮＵＬＬ＃を示し、それは第１のパイプのライニングが取り替えられていないことを示し得る。各パイプのライニングが取り替えられているかどうかを知ることで、パイプ漏れ予測システムは、パイプ漏れの可能性に対するパイプのライニングに取り替えの影響を判断し得る。

【００３７】

いくつかの実施形態において、データは、列３１８（ヘッダ「plr_num」で示される）および列３２０（ヘッダ「plr_year」で示される）を含むことができ、後者は、パイプが設置される年の数値変数を意味し得る。例えば、列３２０は、第１のパイプに対して１９３４を示し、それは第１のパイプが１９３４年に設置されたことを示し得る。各パイプが設置された年を知ることにより、パイプ漏れ予測システムは、パイプ漏れの可能性に対する設置年の影響（例：おそらく数年間パイプが不適切に設置され、漏れる可能性がより高い）を判断し得る。

【００３８】

いくつかの実施形態では、データは、パイプが設置される土壌の抵抗率についての数値変数を示す列３２２（ヘッダ「soil_res」で示される）を含んでもよい。つまり、この列は、パイプが設置されている土壌の腐食性を示す。腐食によりパイプが腐食するため、腐食性土壌にある新品のパイプの方が、非腐食性土壌にある古いパイプよりも早く漏れることが予想され得る。より高い値の土壌抵抗率ほど、より低い腐食性に関連付けられ得る。例えば、列３２２は、第１のパイプの土壌抵抗率について１１３８を示し、第２のパイプの土壌抵抗率について３６０１を示し、これは第２のパイプがより腐食性の低い土壌に位置することを示し得る。各パイプが位置する土壌の相対的な腐食性を知ることにより、パイプ漏れ予測システムは、パイプの漏れの可能性に対する土壌腐食性の影響を判断し得る。

【００３９】

一部の実施形態では、データは、パイプが漏れた回数についての数値変数を示す列３２４（ヘッダ「leak_freq」で示される）を含むことができる。つまり、値０は、パイプがまだ漏れしていないことを示す。図に示されるように、列３２４は、第１のパイプの漏れ頻度について０を示し、これは、第１のパイプがまだ漏れていないことを示し得る。これは、パイプが漏れたかどうかのカテゴリ変数（たとえば、「Ｎ」または「Ｙ」の値）を示す列３２６（ヘッダ「leak_yn」で示される）に適切に対応する。列３２６は、第１のパイプについて、列３２４の０の値に適切に対応するＮの値を有し、第１のパイプがまだ漏れていないことを示す。場合によっては、以前に漏れたパイプが再び漏れる可能性が高くなったかもしれない。パイプが以前に漏れたかどうかを知ることにより、パイプ漏れ予測システムは、パイプ漏れの可能性に対する以前の漏れの影響を判断し得る。

【００４０】

いくつかの実施形態では、データは、パイプが地理的にどこに位置するかを示すカテゴリ変数を示す列３２８（ヘッダ「district（地区）」で示される）を含むことができる。変数の例には、East Valley（東谷地区）、Western（西部地区）、West Valley（西谷地区）、Harbor（港湾地区）、およびCentral（中央地区）が含まれる。地理的位置の、他

の変数または概念が用いられてもよい。図では、列３２８は、第１のパイプがどこにあるかを示すために、第１のパイプについて、East Valleyの値を有する。パイプの位置は、（たとえば、その地理的位置の土壌の構成により）パイプの漏れの可能性に何らかの影響を与える可能性があり、したがって、この情報により、パイプ漏れ予測システムはその影響を判断できる。

【００４１】

いくつかの実施形態では、データは、パイプの年齢についての数値変数を意味する列３３０（ヘッダ「pipe_age（パイプ年齢）」で示される）を含むことができる。例えば、列３３０は、第１のパイプについて、それが８２年の年齢であることを示し得る値「８２」を有する。

【００４２】

いくつかの実施形態において、データは、パイプの材料の寿命に関する数値変数を示す列３３２（ヘッダ「material_life（材料寿命）」で示される）を含み得る。例えば、列３３０は第１のパイプについて値「１００」を有し、これは、鋼で形成される第４のパイプの１２０年の材料寿命と比較して、第１のパイプを構成する鉄が１００年の寿命を有することを示し得る。

【００４３】

いくつかの実施形態では、データは、パイプが製造業者の仕様に対してどれだけ古いかを示す数値変数を示す列３３４（ヘッダ「pct_life」で示される）を含むことができる。いくつかの実施形態では、パイプの百分率寿命は、パイプが材料寿命（たとえば、列３３２）に対してどれくらい古いか（たとえば、列３３０）を判定することによって計算され得る。たとえば、第１のパイプの年齢は８２年で、材料寿命は１００年であり、その結果、列３３４に示される値は８２％である。パイプのその寿命に対する古さの、この正規化されたメトリックは、パイプ漏れ予測システムによって使用されて、パイプの相対年齢がそのパイプ漏れの可能性にどのように影響し得るかを判断し得る。

【００４４】

いくつかの実施形態では、データは、パイプが耐えることができる最大圧力についての数値変数を示す列３３６（ヘッダ「max_pressure（最大圧力）」で示される）を含むことができる。任意の測定単位を事前に指定して用いることができる。図においては、列３３６の数字は、パイプが１平方インチあたりポンド（ｐｓｉ）で耐えることができる圧力を表してもよい。たとえば、列３３６は、第１のパイプについて３３３を示し、それは第１のパイプが３３３ｐｓｉの最大圧力を定格とされていることを示し得る。パイプ漏れ予測システムは、この情報を用いて、パイプ漏れの可能性に対する最大圧力定格の影響を判断し得る。

【００４５】

いくつかの実施形態では、データは、パイプの標高についての数値変数を示す列３３８（ヘッダ「pipe_elevation（パイプ標高）」で示される）を含んでもよい。任意の測定単位を事前に指定して用いることができる。図において、列３３８の数字は、パイプの標高をフィート単位で表してもよい。例えば、列３３８は、第１のパイプに対して６６６を示し、それは、第１のパイプが海抜６６６フィートに位置することを示し得る。パイプ漏れ予測システムは、この情報を用いて、パイプの標高がパイプ漏れの可能性に及ぼす影響を判断し得る。

【００４６】

いくつかの実施形態では、データは、パイプが丘陵地にあるかどうかを示すカテゴリ変数（例えば、「Ｎ」または「Ｙ」の値）を示す列３４０（ヘッダ「within_hill（丘陵地内）」で示される）を含み得る。例えば、列３４０は、第１のパイプに対してＮを示し、それは、第１のパイプが丘陵地に位置していないことを示し得る。直感的には、丘陵地にあるパイプは、丘陵地からパイプにかかる追加の圧力のため、漏れる可能性がより高くなり得る。パイプ漏れ予測システムは、この確認を用いて、その直感を確認し、パイプが丘陵地にあるかどうかがパイプ漏れの可能性に与える影響を判断し得る。

【００４７】
図４は、本開示の実施形態によるパイプデータベース内のパイプレコードの例示的な数字を示す。
【００４８】
より具体的には、図４は、２６５，２９３本のパイプに関連する情報を含むパイプデータベース４０２を示す。これらのパイプのうち、２５６，８０３本には漏れがなく、８，４９０本のパイプには漏れがある。パイプデータベース４０２内のこのデータは、パイプ漏れ予測システムを訓練して漏れを検出するために使用でき、本明細書で開示される例示的な実施形態は、これら２６５，２９３本のパイプの場合において説明されるさまざまな機械学習モデルを利用し得る。
【００４９】
図５は、本開示の実施形態によるランダムフォレストモデルの混同行列を示す。
いくつかの実施形態では、さまざまなパイプに関連付けられる利用可能なデータのすべて（例えば、図３に示されるものと同様のデータ）を用いて、予測モデルを訓練および構築することができる。そのような実施形態のいくつかでは、構築される第１のモデルは、非線形モデルであるランダムフォレストモデルであり得る。一般に、非線形モデルは、線形モデルよりも訓練に時間がかかり得る。
【００５０】
いくつかの実施形態では、ランダムフォレストモデルを用いて、図３に示される因子またはパイプ属性などの１１個の因子の組み合わせに基づいてパイプが漏れるかどうかを予測し得る。これらの１１の因子から除外されるのは、漏れ頻度に関連付けられる入力変数（例えば、図３の「leak_freq」変数）であり得る。これは意図的であり、なぜならば、既存の専門家ベースのモデルは、パイプが漏れたかどうかの過去の知識が、パイプが将来漏れるかどうかを予測する上で最も重要な因子である、という仮定に基づき得るからである。しかしながら、この仮定は、予測された漏れが、パイプが以前に漏れたかどうかに大きく依存することを意味し、過去に漏れたことがないパイプが漏れるかどうかを予測することを困難なものにする（たとえば、問題をパイプ予報問題に変換する）。対照的に、漏れ頻度の入力を除外することにより、ランダムフォレストモデルは、パイプが実際に漏れたという事前の知識なしに、パイプが漏れるかどうかを予測するように構成され得る。
【００５１】
ランダムフォレストモデルは、二値結果の予測をもたらす分類ベースのモデルであり得る。言い換えれば、ランダムフォレストモデルは、ラベル付けされたカテゴリカルな予測（例：パイプが漏れるかどうかについての「はい」または「いいえ」）を生成するように訓練できる。ランダムフォレスト分類の出力は、フォレスト内の個々のツリーの多数決であるため、これらのカテゴリカルな予測は、出力がカットオフまたはしきい値が特定されることをさらに必要とする確率である回帰手法（ロジスティック回帰など）の結果とは異なる。ランダムフォレストモデルは、パイプデータベースからの訓練用セットを用いて訓練できる。たとえば、さまざまなパイプについての情報を含むパイプデータベースは、訓練用セットとテスト／検証セットとにランダムに分割でき、７０／３０または８０／２０分割が頻繁に用いられる。ランダムフォレストモデルは、訓練用セットを用いて訓練できる。
【００５２】
その後、ランダムフォレストモデルは、パイプラインデータベースからの検証セットを用いて検証およびテストできる。ランダムフォレストモデルを、検証セット内のパイプについてのさまざまな入力とともに用いて、それらのパイプが漏れるかどうかを予測できる。ランダムフォレストモデルからの予測結果は、それらのパイプが漏れたかどうかの検証セットからの実際の知識と比較され得る。その後、混同行列を作成して、パイプについての予測の正答率を、それらのパイプの真のステータスに対して評価し得る。
【００５３】
図に関して、図４に示されたデータ（例：２６５，２９３本のパイプを含むデータセット）に基づいて生成されたランダムフォレストモデルについて、例示的な混同行列５０２が

図５に示される。混同行列５０２は、予測を実際の参照値と比較するために４つの象限：２５６，８０３の値を有する左上のＮ－Ｎ象限、１０７の値を有する右上のＮ－Ｙ象限；０の値を有する左下のＹ－Ｎ象限；８３８３の値を有する右下のＹ－Ｙ象限を有する。混同行列５０２のこれらの象限は、４つの結果：真陰性、偽陰性、真陽性、および偽陽性の数を反映している。

【００５４】

２５６，８０３の値を有する左上のＮ－Ｎ象限は、真陰性、つまり、モデルが故障しないと予測し、実際に故障しなかったパイプの数を表す。１０７の値を有する右上のＮ－Ｙ象限は、偽陰性、つまり、モデルが故障しないと予測したが、実際には故障したパイプの数を表す。０の値を有する左下のＹ－Ｎ象限は、偽陽性、つまり、モデルが故障するであろうと予測したが、実際には故障しなかったパイプの数を表す。８３８３の値を有する右下のＹ－Ｙ象限は、真陽性、つまり、モデルが故障するであろうと予測し、実際に故障したパイプの数を表す。

【００５５】

モデルの全体的な正答率５０４は高く、約９９．６％である。感度５０６とも呼ばれるモデルの真陽性率は９８．７％であり、モデルが８，４９０のパイプ漏れのうち８，３８３（または９８．７％）を検出したことに対応する。したがって、「パイプが漏れるか」という質問でプロンプトが表示されると、モデルは９８．７％の正確な回答を返すことになる。特異性５０８とも呼ばれるモデルの真陰性率は１００％であり、これは、モデルが２５６，８０３本のパイプのうち故障しなかったすべてを正しく判断したことに対応する。したがって、「パイプは漏れないか」という質問でプロンプトが表示されると、モデルは１００％正確な回答を返すことになる。したがって、混同行列に示されたこれらの検証結果に基づくと、偽陽性率および偽陰性率は非常に低く、構築されたランダムフォレストモデルは、漏れるパイプおよび漏れないパイプを正確に識別する能力が高いことがわかる。

【００５６】

図６は、本開示の実施形態によるランダムフォレストモデルの変数重要度プロットを示す。

【００５７】

図５に記載されたランダムフォレストモデルは、パイプ漏れの判定に用いられるさまざまな変数および因子の重要度を判定するためにも用いることができる。さまざまな変数の相対的な重要度を変数重要度プロットでプロットし視覚化できる。

【００５８】

図に関して、ランダムフォレストモデルの生成に用いられる１１の因子の重要度を示す変数重要度プロット６００が図６に示される。変数重要度プロット６００は、これらの因子の中で最も重要なものがパイプの長さ、土壌抵抗率、パイプの最大圧力定格、およびパイプの標高であることを示す。特に、パイプの長さおよび土壌の特性は、パイプが丘陵地にあるか、またはパイプがどのように使用されているかよりも、パイプの漏れを判断する上で有意により重要である。実際、これらの４つの変数は非常に重要であるため、ランダムフォレストモデルは、これら４つの変数についてのデータを用いるだけで、パイプが漏れるかどうかについて非常に正確な予測を生成し得る。この情報は非常に有用であり、なぜならば、それは、データ収集に関連して多大なコストがかかる（たとえば、さまざまな因子についての値を各パイプごとに取得するのは困難であるかまたはコストがかかる）場合に、これら４つの変数の値だけを収集することができるので、満足のゆく予測モデルを安価に実現できることを意味するからである。実際、パイプの長さを知るだけで、パイプの漏れに関する比較的正確な予測ができる場合がある。

【００５９】

さらに、パイプ漏れの背後にある最も重要な因子に関するこの知識を用いて、パイプの将来の設置を管理し、将来の漏れを防ぐことができる。この知識を用いて、漏れの可能性がより低いパイプを製造業者または供給業者から取得できる。さらに、この情報を用いて、建設入札を勝ち取りたい製造業者が従わなければならないような新たな製造業者要件／基準を設計できる。たとえば、ランダムフォレストモデルでは、パイプが漏れるかどうかに

影響することについて、パイプの長さが最も重要な因子であるため、将来のパイプの設置では、長いパイプを避けることができる。これらの長いパイプは、各々漏れが発生する可能性がより低い一連の短いパイプに置き換えることができる。また、変数重要度プロット６００によって特定された因子の重要度に関するこの知識は、基礎となるデータに根ざしている。たとえば、腐食性土壌における新品のパイプは、非腐食性土壌における古いパイプよりも寿命がはるかに短くなることが経験的に検証されているため、可能であれば腐食性土壌のない場所でパイプを設置するべきである。

【００６０】

変数重要度プロット６００は、ランダムフォレストモデルの各因子がどれほど重要であるかを示すが、各因子が具体的に重要である理由を説明していない。それを理解するために、これらの推論を可能にする別のモデルが用いられ得る。場合によっては、機械学習技術となると、予測と推論との間にトレードオフが存在する場合がある。ランダムフォレストモデルは、パイプ漏れの正確な予測を作成するのに優れているが、モデルで用いられる因子の関係に関して推論を行うことは可能にしない。したがって、異なるモデルを用いてこれらの推論を生成し得る。用いることができるさまざまなモデルの例には、ロジスティック回帰モデルまたは単純ベイズモデルが含まれる。

【００６１】

図７は、本開示の実施形態によるロジスティック回帰の確率の統計的分布を示す。

いくつかの実施形態では、ランダムフォレストモデルで用いられる因子をよりよく理解するために、ロジスティック回帰モデルを用いることができる。ロジスティック回帰モデルは、一般化線形モデルの一種である。ロジスティック回帰モデルは線形モデルであるため、ランダムフォレストモデルよりも速く訓練できる。

【００６２】

しかしながら、ロジスティック回帰モデルは、各パイプをカテゴリ（例：漏れありまたは漏れなし）に分類するのではなく、連続応答値（パイプ漏れの確率など）を生成する点で、ランダムフォレストモデルとは異なる。したがって、ロジスティック回帰モデルを用いて各パイプごとに同様の分類を行うには、カットオフまたはしきい値を指定する必要がある。ロジスティック回帰モデルによって予測されるパイプ漏れの確率がそのしきい値を上回る場合、パイプが漏れそうであると考えることができる。パイプ漏れの確率がそのしきい値を下回ると予測される場合、パイプは漏れそうでないと考えることができる。このように、そのしきい値を超える確率はｙｅｓになり、そのしきい値を下回る確率はｎｏになるため、適切なしきい値を選択する必要がある。

【００６３】

適切なカットオフを選択するために、ロジスティック回帰モデルの結果を統計分布にプロットし得る。この統計分布は、ロジスティック回帰モデルによって予測される確率を２つのグループに分割するための自然なカットオフを判断するために用いることができる。

【００６４】

図に関して、ロジスティック回帰モデルから生成された確率の統計分布７００が図７に示される。たとえば、統計分布７００では、ｙｅｓ変数の第３四分位数は０．０３レベルの水平の黒いバーである。これがカットオフとして選択された場合、それは、この第３四分位数（０．０３）以上で予測される故障の確率はすべて「ｙｅｓ」と見なされ、パイプは漏れることになることを意味するであろう。このカットオフより下で予測される故障の確率は、「ｎｏ」と考えられるであろう。

図８は、本開示の実施形態によるロジスティック回帰の混同行列を示す。

【００６５】

ロジスティック回帰モデルのカットオフが計算されると、ロジスティック回帰モデルの混同行列を構築するために、ロジスティック回帰モデルの予測結果を（たとえば、以前どおりの検証セットを用いて）検証し得る。

【００６６】

図に関して、図８に、図４に示されたデータ（例：２６５，２９３本のパイプを含むデー

タセット）に基づいて生成されたロジスティック回帰モデルについて、例示的な混同行列８０２が示される。混同行列８０２は、予測を実際の参照値と比較するために４つの象限：１９６，５８３の値を有する左上のＮ－Ｎ象限、２１２３の値を有する右上のＮ－Ｙ象限；６０２２０の値を有する左下のＹ－Ｎ象限；６３６７の値を有する右下のＹ－Ｙ象限を有する。図５の混同行列と同様に、混同行列８０２のこれらの象限は、４つの結果：真陰性、偽陰性、真陽性、および偽陽性の数を反映する。

【００６７】

１９６，５８３の値を有する左上のＮ－Ｎ象限は、真陰性、つまり、モデルが故障しないと予測し、実際に故障しなかったパイプの数を表す。２１２３の値を有する右上のＮ－Ｙ象限は、偽陰性、つまり、モデルが故障しないと予測したが、実際には故障したパイプの数を表す。６０２２０の値を有する左下のＹ－Ｎ象限は、偽陽性、つまり、モデルが故障するであろうと予測したが、実際には故障しなかったパイプの数を表す。６３６７の値を有する右下のＹ－Ｙ象限は、真陽性、つまり、モデルが故障するであろうと予測し、実際に故障したパイプの数を表す。

【００６８】

ロジスティック回帰モデルは、ランダムフォレストモデルに比べて正確ではないことがわかる。ロジスティック回帰モデルの全体的な正答率８０４は約７６．５％である。ロジスティック回帰モデルの偽陽性率および偽陰性率もそれほど大きくなく、ランダムフォレストモデルと比較して、多くの偽陽性および偽陰性がある。これは、ロジスティック回帰モデルが線形モデルであるため、パイプ漏れが線形プロセスではないことを示唆している可能性がある。さらに、感度８０６とも呼ばれるモデルの真陽性率は７４．９％であり、これはモデルが８，４９０のパイプ漏れのうち６，３６７（または７４．９％）を検出したことに対応する。したがって、「パイプが漏れるか」という質問でプロンプトが表示されると、モデルは７４．９％の正確な回答を返すことになる。特異度８０８とも呼ばれるモデルの真陰性率は７６．５％であり、これは、モデルが漏れなかった２５６，８０３本のパイプのうち１９６，５８３本を正確に判断したことに対応する。したがって、「パイプは漏れないか」という質問でプロンプトが表示されると、モデルは７６．５％の正確な回答を返すことになる。

【００６９】

したがって、混同行列で示されたこれらの検証結果に基づいて、ロジスティック回帰モデルは、パイプ漏れを予測することについて、ランダムフォレストモデルほど正確ではないことがわかる。特に、これは、パイプ漏れデータに非線形性があることを示している可能性があり、なぜならば、基礎データが線形である場合、線形モデルは非線形モデルよりも優れているためである。しかしながら、これで問題ない場合があり；なぜならば、ロジスティック回帰モデルは、推論を可能にし、入力変数間の関係を判断するのに役立ち得、それは、ランダムフォレストモデルからは得られない視点を提供できるからである。たとえば、ロジスティック回帰モデルを用いて、ロジスティック回帰モデルにおける各因子ごとの回帰係数および各因子の統計的有意性を取得できる。

【００７０】

図９は、本開示の実施形態によるロジスティック回帰の係数のリストを示す。

より具体的には、ロジスティック回帰モデルに対する係数のリスト９０２を用いて、ロジスティック回帰モデルで用いられるさまざまな因子について推論を行うことができる。横に星印の付いた係数は統計的に有意なものであり、偶然に発生した可能性は低いことを意味する。たとえば、パイプの直径、パイプの長さ、パイプの厚み、最大圧力定格、パイプがＨＤＰＥ製かＰＶＣ製か、パイプのパーセント寿命などの、パイプに基づく因子は、統計的に有意であるとして示される。土壌の抵抗率、パイプの地理的領域、パイプの標高、パイプが丘陵地にあるかどうかなどの、場所に基づく因子も、統計的に有意であるとして示される。

【００７１】

パイプの長さについては、標準誤差は、関連するｐ値がゼロに近い非常に小さな数値であ

ることを通知する。より具体的には、パイプ長は９９パーセンタイルレベルで因子として統計的に有意である。さらに、係数は正であり、これは、パイプの長さが長くなると、パイプが故障する可能性が高くなることに関連付けられることを意味する。ランダムフォレストモデルはパイプの長さが非常に重要であると通知したが、理由を特定しなかったため、これは有用な情報である。ランダムフォレストモデルとロジスティック回帰モデルとの両方が同じポイントで一致する場合、より具体的な枠組みが取得される。パイプの長さに関するこの情報は、複数のパイプを購入する必要があることを意味する場合でも、より短いパイプを購入すべきであることを意味する。これにより初期コストが増大する可能性があるが、パイプの故障の可能性が大幅に減少するであろう。

【００７２】

パイプの直径については、係数は、パイプの直径がパイプの故障と負の関連があることを示す。これは、パイプの直径が大きくなると、パイプが故障する可能性がより低くなることを意味する。つまり、幅の広いパイプは故障する可能性がより低く、細いパイプは故障する可能性がより高くなる。これは有用な情報であり、なぜならば、幅の広いパイプは初期費用がより高くなる可能性があるが、パイプが故障する可能性がより低い場合、長期的にはより安価になる可能性があるからである。パイプの厚みについては、係数は負であり、これは、より厚い壁部のパイプは、より薄いパイプよりも故障しにくいことを意味する。この情報は有用な場合があり、なぜならば、より厚い壁部のパイプはより高価だが、故障する可能性はより低いため、より厚い壁部のパイプを使用するよう、より多くの先行投資を行う方が合理的であり得るからである。

【００７３】

パイプの寿命の割合も重要である。パイプが製造業者の推奨する寿命を超えると、故障する可能性がより高くなる。これは常識であり、直感を裏付ける。しかしながら、驚くべきことに、パイプがどのように用いられるかは統計的に重要ではない。これは、ランダムフォレスト変数重要度プロットと一致するようであり、この一致は両方のモデルから得られる洞察を裏付ける。これらの説明された推論は、引き出され得る推論の網羅的なリストではなく、むしろそれらのほんの一部である。２つのモデルから多数の推論を直接引き出すことができ、具体的かつ数学的に厳密な態様で以前の仮定を確認または検証するために用いることができる貴重な情報を提供する。

【００７４】

追加の洞察を得るために、追加の機械学習技術またはモデルを用いることができる。たとえば、機械学習モデルの組み合わせには、推論の生成にも用いることができる単純ベイズモデルが含まれ得る。

【００７５】

図１０は、本開示の実施形態による単純ベイズモデルの混同行列を示す。

図に関して、図４に示されたデータ（例:２６５，２９３本のパイプを含むデータセット）に基づいて生成された単純ベイズモデルについて、例示的な混同行列１００２が示される。混同行列１００２は、予測を実際の参照値と比較するために４つの象限:２４４，９９５の値を有する左上のＮ－Ｎ象限、５７３１の値を有する右上のＮ－Ｙ象限;１１８０８の値を有する左下のＹ－Ｎ象限;２７５９の値を有する右下のＹ－Ｙ象限を有する。図５の混同行列と同様に、混同行列１００２のこれらの象限は、４つの結果：真陰性、偽陰性、真陽性、および偽陽性の数を反映する。

【００７６】

２４４，９９５の値を有する左上のＮ－Ｎ象限は、真陰性、つまり、モデルが故障しないと予測し、実際に故障しなかったパイプの数を表す。５７３１の値を有する右上のＮ－Ｙ象限は、偽陰性、つまり、モデルが故障しないと予測したが、実際には故障したパイプの数を表す。１１，８０８の値を有する左下のＹ－Ｎ象限は、偽陽性、つまり、モデルが故障するであろうと予測したが、実際には故障しなかったパイプの数を表す。２７５９の値を有する右下のＹ－Ｙ象限は、真陽性、つまり、モデルが故障するであろうと予測し、実際に故障したパイプの数を表す。

【００７７】
単純ベイズモデルは、ランダムフォレストモデルに比べて正確ではないことがわかる。単純ベイズモデルの全体的な正答率１００４は約９３．４％である。単純ベイズモデルの偽陽性率および偽陰性率もそれほど大きくなく、ランダムフォレストモデルと比較して、多くの偽陽性および偽陰性がある。これは、単純ベイズモデルが線形モデルであるため、パイプ漏れが線形プロセスではないことを示唆している可能性がある。さらに、感度１００６とも呼ばれるモデルの真陽性率は３２．５％であり、これは、モデルが８，４９０のパイプ漏れのうちわずか２７５９（または３２．５％）を検出したことに対応する。したがって、「パイプが漏れるか」という質問でプロンプトが表示されると、モデルは３２．５％の正確な回答を返すことになる。特異度１００８とも呼ばれるモデルの真陰性率は９５．４％であり、これは、モデルが漏れのない２５６，８０３本のパイプのうち２４４，９９５本のパイプを正しく判断したことに対応する。したがって、「パイプは漏れないか」という質問でプロンプトが表示されると、モデルは９５．４％の正確な回答を返すことになる。したがって、単純ベイズモデルを用いて、パイプが漏れない場合を妥当な正答率で判断できるが、パイプが漏れるかどうかを判断するのには用いるべきではない。
【００７８】
典型的には、単純ベイズモデルは、通常非常に高速に訓練されるため、最適なモデルではなく、ベースラインとして用いられ得る。これは、１１の入力（ロジスティック回帰およびランダムフォレストで用いられる同じ入力）の各々が互いに独立しているという、単純ベイズモデルで用いられる仮定のためである。これが当てはまる場合、単純ベイズモデルの結果は非常に正確であることになる。この仮定が当てはまらない場合、モデルの結果は正確ではないことになる。したがって、単純ベイズモデルの正確さまたは不正確さを確認することで、入力が相互にどの程度依存しているかについての推論を引き出すことができ、これは、ランダムフォレストモデルおよびロジスティック回帰モデルからの洞察と組み合わせた場合、有用な情報である。これらの結果は、ロジスティック回帰モデルとランダムフォレストモデルとの両方を下回っており、これは、因子変数の独立性（単純ベイズモデルの仮定）が有効ではなさそうであること、およびこれらの予測変数自体の間に、モデルの予測正答率に影響を与えるのに十分重要な複雑な相互関係があることを示唆している。この洞察は、各因子に個別に重みを割り当てるモデルは不正確でありそうであることを意味する。
【００７９】
図１１は、本開示の実施形態に従った、パイプ漏れ予測を実現するためのブロック図である。
【００８０】
いくつかの実施形態においては、パイプ漏れ予測システムについての解決実現例はビッグデータクラウドプラットフォーム１１１０を含み得る。いくつかの実施形態においては、ビッグデータクラウドプラットフォーム１１１０は、パイプデータ（例えばパイプ特性および既知の漏れ）をすべて、標準フォーマット１１２２でオブジェクトストア１１２０に格納し得る。
【００８１】
いくつかの実施形態においては、オブジェクトストア１１２０が維持されて、ビッグデータクラウドサービス１１４０を介して提供されてもよく、それはアマゾンシンプルストレージサービス（Ｓ３）のようなクラウドコンピューティングサービスであることができる。いくつかの実施形態においては、ビッグデータクラウドサービス１１４０は、分散型ストレージおよび大型のデータセットの処理のために用いられるオープンソースソフトウェアフレームワークであるApache Hadoopを用いて実現されてもよい。ビッグデータクラウドサービス１１４０は、オブジェクトストア１１２０におけるデータのすべてを格納するよう用いられてもよい。したがって、オブジェクトストア１１２０に含まれるデータは分割されて、ビッグデータクラウドサービス１１４０のコンピューティングクラスタにおけるノードにわたって分散された大型のブロックに格納され得る。いくつかの実施形態に

おいては、ビッグデータクラウドサービス１１４０は、スケーラブルで信頼性の高いデータストレージを提供するために使用されるJava（登録商標）ベースのファイルシステムであるHDFS（Hadoop分散型ファイルシステム）１１４６を使用して実装できる。

【００８２】

いくつかの実施形態では、ビッグデータクラウドサービス１１４０は、データのストリーミングおよび機械学習のために組み込みモジュールを介してビッグデータ処理用の高速かつ一般的なエンジンとして機能するクラスタコンピューティングフレームワークを提供し得るＳｐａｒｋ１１４４（例えば、Apache Spark）を利用し得る。言い換えると、Ｓｐａｒｋ１１４４は、コンピュータのクラスタにわたる大規模なデータセットの分散型処理を用いてビッグデータ分析を実行するための機能を提供し得、大量のデータの分散型処理のための基本hadoopmap/reduce技法上で改善し得る。Ｓｐａｒｋ１１４４はメモリに永続化することでネイティブのHadoopMap/Reduce機能を大幅に強化し得、一方Map/Reduceはディスクに永続化される。その結果、Ｓｐａｒｋ１１４４は、メモリ動作の点でMap/Reduceの１００倍以上、ディスク動作の点でMap/Reduceの１０倍以上高速であることができる。

【００８３】

図に示されるように、ビッグデータクラウドプラットフォーム１１１０は、バルクソースデータ１１５０およびストリーミングソースデータ１１５２を受信し得る。一般に、バルクソースデータ１１５０は、現場に出ているパイプの特性、およびそれらのうちのどれが漏れたか、ならびに、もはや現場にないパイプの特性、およびそれらのうちのどれが漏れたかなどの履歴データを含み得る。いくつかの実施形態においては、バルクソースデータ１１５０は、「データレイク（data lake）」と称され得るオブジェクトストア１１２０に格納されてもよい。オブジェクトストア１１２０に格納されたデータはすべて、一様に、標準フォーマット１１２２で格納されてもよい。いくつかの実施形態においては、標準フォーマット１１２は、スケーラブルで信頼性の高いデータストレージを提供するために用いられるJavaベースのファイルシステムであるＨａｄｏｏｐ分散型ファイルシステム（Hadoop Distributed File System：ＨＤＦＳ）であってもよい。

【００８４】

ストリーミングソースデータ１１５２は、現場に配置された新たなパイプについてのデータ（たとえば、それらのパイプの特性）、およびどのパイプが漏れたかの更新を含むことができる。このデータは、（たとえば、新たなパイプを設置したり、既存のパイプの漏れをチェックしている現場スタッフから）リアルタイムで受信され得る。いくつかの実施形態では、現場スタッフは、このデータを現場でコンピューティングデバイスまたはモバイルデバイス上で入力することができ、そのデータはビッグデータクラウドプラットフォーム１１１０によって受信され、オブジェクトストア１１２０に統合されることになる。いくつかの実施形態においては、ストリーミングソースデータ１１５２は、Ｋａｆｋａ１１３０などの通信インターフェイスにおいて受信されてもよい。通信インターフェイスとして、Ｋａｆｋａ１１３０はストリーミングソースデータ１１５２を受信し、そのデータ内の個々のメッセージを解析してもよい。それらのメッセージは、オブジェクトストア１１２０に（たとえば、標準フォーマット１１２２で）格納することができるデータに変換することができる。したがって、オブジェクトストア１１２０は、バルクソースデータ１１５０（例えば、既存のパイプの履歴特性）およびストリーミングソースデータ１１５２（例えば、新たなパイプの特性または既存のパイプの履歴特性への更新）からのデータを同じ一様なフォーマットに含むことができる。

【００８５】

いくつかの実施形態においては、Ｓｐａｒｋ１１３４および／またはＨｉｖｅ　ＬＬＡＰ１１３６はさらに、オブジェクトストア１１２０内に含まれているデータをすべて（たとえば不正な挙動を検出するために）分析して処理するために用いられる。Ｓｐａｒｋ１１３４（たとえば、Ａｐａｃｈｅ　Ｓｐａｒｋ）は、データのストリーミングおよび機械学習のために内蔵モジュールを介してビッグデータ処理のための高速かつ汎用のエンジンと

して機能するクラスタコンピューティングフレームワークを提供し得る。言いかえれば、Ｓｐａｒｋ１１３４は、コンピュータのクラスタにわたる大規模なデータセットの分散処理を用いてビッグデータ分析を実行するための特徴を提供し得るとともに、大量のデータを分散処理するためのベースとなるＨａｄｏｏｐ　Ｍａｐ／Ｒｅｄｕｃｅ技術を改善させ得る。Ｓｐａｒｋ１１３４は、メモリに存続することによって固有のＨａｄｏｏｐ　Ｍａｐ／Ｒｅｄｕｃｅ機能を過剰にチャージし得る一方で、Ｍａｐ／Ｒｅｄｕｃｅはディスクに残存している。結果として、Ｓｐａｒｋ１１３４は、メモリ動作の点でＭａｐ／Ｒｅｄｕｃｅよりも１００倍高速であり得るとともに、ディスク動作の点でＭａｐ／Ｒｅｄｕｃｅよりも１０倍高速であり得る。

【００８６】

いくつかの実施形態においては、Ｓｐａｒｋ１１３４は、Oracle R Advanced Analytics for Hadoop（ＯＲＡＡＨ）を含み得る。Oracle R Advanced Analytics for Hadoopは、コンピュータのクラスタにわたる大型のデータセットの分散処理を用いて、ビッグデータ分析を実行するための特徴を提供する「スーパーチャージされた（supercharged）」バージョンのＳｐａｒｋとしての役割を果たし得る。ＯＲＡＡＨは、従来のＳｐａｒｋパッケージに勝る多数の利点を提供し得る。たとえば、ＯＲＡＡＨは、スパークより３２倍高速の機械学習モデル（たとえば、分類、クラスタ化、回帰、特徴抽出などのための機械学習アルゴリズム）を提供し得る。ＯＲＡＡＨはまた、研究開発において開発された機械学習モデルを製造に向けて展開する能力を提供してもよい。ＯＲＡＡＨはまた、データレイクにおいてＲスクリプトを直接実行する能力を提供してもよい。ＯＲＡＡＨはまた、ＨＤＦＳおよび／またはＨＩＶＥを含む複数のデータフォーマットからのデータの読取り／書込みを可能にする単一のパッケージとしての役割を果たし得る。ＯＲＡＡＨはまた、Ｒ内に存在する任意の式を処理可能であり得るのに対して、Ｓｐａｒｋは、制限された変換のサブセットで単純な属性を処理することができるだけであり得る。

【００８７】

いくつかの実施形態においては、Ｈｉｖｅ　ＬＬＡＰ１１３６は、Ａｐａｃｈｅ　Ｈｉｖｅを含んでいてもよく、データ要約、クエリおよび分析のためにＳＱＬ様のインターフェイスを提供するためのＡｐａｃｈｅ　Ｈａｄｏｏｐプラットフォーム上に構築されたデータウェアハウスソフトウェアプロジェクトであり得る。Ｈｉｖｅ　ＬＬＡＰ（低レイテンシ分析処理）１１３６は、より高速のＳＱＬ分析を提供することによってＨｉｖｅアーキテクチャ上に構築されてもよい。これにより、Ｋａｆｋａ１１３０を用いて、ストリーミングデータからのメッセージを解析して取得する。これらメッセージは次いで、オブジェクトストア１１２０における履歴データに追加される。オブジェクトストア１１２０に含まれるこの「データレイク」は、この明細書中に上述されてきた電力不正を検出するためのステップを実行するために、Ｓｐａｒｋ１１３４およびＨｉｖｅ　ＬＬＡＰ１１３６を用いて処理される。

【００８８】

いくつかの実施形態においては、オブジェクトストア１１２０は、コンピューティング効率の向上および必要なコンピューティングリソースの減少に関連付けられた特徴を、分離されたストレージに提供する態様で実現されてもよい。Ｈａｄｏｏｐは、典型的には、ＨＤＦＳとＭａｐＲｅｄｕｃｅとの組合せから成る。しかしながら、ＨＤＦＳに関する問題は、計算が各ノード（たとえば、分散型コンピューティングシステムのクラスタ）上にあって、付加的な計算を得るためにより多くのノードを追加する必要がある点である。各々のノードは計算および格納を行なうものであって、これは、より多くのノードを追加することによって、使用されていないストレージが有効に補償されていることを意味している。代替例として、ＨＤＦＳ以外のストレージ機構、たとえばＡｍａｚｏｎ　Ｓ３またはオラクルオブジェクトストレージなどを用いることができる。たとえば、システムがオブジェクトストレージとＭａｐＲｅｄｕｃｅとの組合せで実現されるように、ＨＤＦＳを交換することができる。この実現例の下では、ストレージは分離されており、ノードには最小限のストレージを追加することができ、そのストレージに関連付けられ得る追加コストを

低減することができる。言いかえればオラクルのビッグデータクラウドサービスなどのサービス（計算エディション）は、追加のＨａｄｏｏｐクラスタまたはＳｐａｒｋクラスタを要求に応じてプロビジョニングするために用いることができるが、データ自体は、Ａｍａｚｏｎ　Ｓ３またはオラクルオブジェクトストレージ内に保持されており、必要に応じてクラスタによって検索される。

【００８９】

図１２は、本開示の実施形態に従った、パイプ漏れ予測のためのハイブリッドシステム図を示す。

【００９０】

図に示すように、パイプ漏れ予測システム１２２０はパイプデータベース１２３４のデータにアクセスできる。このデータベースには、現場に存在するパイプのパイプ特性データ、および過去にどのパイプが漏れたかの知識（既知の漏れなど）が含まれる。このデータは、パイプの漏れを予測できるモデルの構築に役立つ。加えて、パイプ漏れ予測システム１２２０は、パイプラインデータ１２３０にアクセスし得、これには、現場のパイプの新規または更新されたパイプ特性データが含まれ得る。パイプ特性データは、新たに設置されたパイプの場合は新たであり得、場合によっては、（たとえば、現場スタッフ１２１０またはパイプを設置している人によって）新たなパイプ特性データが手動でパイプ漏れ予測システム１２２０に入力され得る。パイプラインデータ１２３０には、現場のパイプとともに、またはその周りに設置されたセンサから取得した環境データも含まれる場合がある。たとえば、パイプラインデータ１２３０には、土壌または標高センサから取得できる土壌データ、標高データ、気象データなどを含めることができる。パイプ漏れ予測システム１２２０は、予測モデルをパイプラインデータ１２３０のデータに適用して、パイプラインデータ１２３０のパイプに関する漏れ予測を出力し得る。

【００９１】

パイプ漏れ予測システム１２２０は、ブロック１２４２においてこのデータをすべて取込んでもよい。これは、現場において任意の環境センサと通信するとともにそれらからデータを受信するように構成された特化された通信インターフェイス（たとえば、プログラミングインターフェイスまたはＡＰＩ）を含んでいてもよい。パイプデータベース１２３４を格納するいずれかのコンピュータシステムまたはデバイスと通信するとともにパイプデータベース１２３４を格納するいずれかのコンピュータシステムまたはデバイスからデータを受信するように構成された通信インターフェイスがあってもよい。たとえば、パイプデータベース１２３４がクラウドコンピューティングネットワーク上で分散された態様で実際に格納される場合、パイプ漏れ予測システム１２２０は、クラウドコンピューティングネットワークからそのデータをすべて検索するための通信インターフェイスを有していてもよい。

【００９２】

データのすべてがパイプ漏れ予測システム１２２０内で統合されると、ブロック１２４４において、パイプ漏れ予測システム１２２０は、データをすべて単数の一様なフォーマットに（たとえば、日付／時間がすべて同じフォーマットに従うことを確実にして）変換し得る。これにより、（たとえば、パイプのデータを入力するさまざまな現場スタッフがさまざまなデータ形式または規約を使用した場合でも）予測モデルを簡単に生成および適用できる。

【００９３】

ブロック１２４６において、パイプ漏れ予測システム１２２０は、データ（たとえば、存在するすべてのパイプについてのパイプ特性）をすべて図１１に示されるオブジェクトストア１１２０などのストレージに格納し得る。ブロック１２４８において、パイプ漏れ予測システム１２２０は、ストレージ内のデータのすべてに対してＲ分析を実行し得る。たとえば、パイプ漏れ予測システム１２２０は、既知のパイプ漏れのデータを用いて、任意のパイプが漏れる可能性に対する、各変数、またはデータが既知のパイプ特性の、相対的な影響を計算し得る。ブロック１２５０で、パイプ漏れ予測システム１２２０は、（例え

ば、一定期間内で）現場内の各パイプが漏れる可能性を予測し、もっとも漏れそうなパイプまたはすでに漏れていそうなパイプを特定し得る。

【００９４】

パイプ漏れ予測システム１２２０がもっとも漏れそうなパイプまたはすでに漏れていそうなパイプを識別すると、エグゼクティブチーム１２９０のメンバーは、デバイス１２８０上のインターフェイス１２８２を介して、それらの識別されたパイプに関するレポートを閲覧することができるようになり得る。デバイス１２８０は、パイプ漏れ予測システム１２２０から生成されたこれらのレポートを受信してもよい。さらに、エグゼクティブチーム１２９０のメンバーは、（たとえば、パイプに関連付けられたＩＤに基づく）所与のパイプについてのパイプ漏れ予測、およびそれらの漏れ予測を生成するために使用される計算の内訳を引き出して閲覧することができてもよい。したがって、エグゼクティブチーム１２９０のメンバーは、漏れ予測をさらに検証および確認し得る。たとえば、メンバーは、非常に古いパイプが漏れそうであると予測されていることに気付き得、それは直感的に理解され得る。そのメンバーは、インターフェイス１２８２内で直接指示して、現場スタッフ１２１０にパイプの場所に物理的に行って漏れがないかパイプを検査するよう指示することができる。

【００９５】

次いで、デバイス１２８０は、パイプ漏れ予測システム１２２０に命令を送信することとなり、さらに、パイプ漏れ予測システム１２２０は、パイプの場所に（たとえば地理的に）最も近くにいる現場スタッフ１２１０を判断することになる。調査が必要なパイプが複数ある場合、これらのパイプは、近くに位置するか否かに基づいて現場スタッフ間で振り分けられることもあり得る（たとえば、現場スタッフ１２１０は、彼らの地理的位置にある、調査のためのパイプのプールを受取ってもよい）。次いで、パイプ漏れ予測システム１２２０は現場スタッフ１２１０に関連付けられたデバイス１２１２に命令を転送し得る。現場スタッフ１２１０は、デバイス１２１２上のインターフェイス１２１４を介して命令およびパイプの場所を閲覧することができてもよい。その後、現場スタッフ１２１０は、パイプの場所を物理的に訪れて、漏れがあるか調査し得る。現場スタッフ１２１０は、デバイス１２１２上のインターフェイス１２１４を介して、パイプがすでに漏れているかどうかを示すことができてもよい。この情報は、エグゼクティブチーム１２９０に報告し返すことができるか、または、パイプ漏れを識別するための任意の既存の予測モデルを更新するかもしくは向上させるために（たとえば、パイプ漏れ予測システム１２２０に格納されている）既存のデータに追加されてもよい。

【００９６】

付加的な実現例の詳細

図１３は、この明細書中に開示された実施形態のうちの１つを実現するための分散型システムを示す簡略図である。分散型システム１３００は、上述したように、パイプ漏れ予測システムの実施形態を実現することができる。例示された実施形態においては、分散型システム１３００は、１つ以上のネットワーク１３１０を介して、ウェブブラウザ、プロプライエタリクライアント（たとえばオラクルフォーム）などのクライアントアプリケーションを実行して動作させるように構成される１つ以上のクライアントコンピューティングデバイス１３０２、１３０４、１３０６および１３０８を含む。サーバ１３１２は、ネットワーク１３１０を介してリモートクライアントコンピューティングデバイス１３０２、１３０４、１３０６および１３０８と通信可能に結合されてもよい。

【００９７】

さまざまな実施形態においては、サーバ１３１２は、システムの構成要素のうち１つ以上によって提供される１つ以上のサービスまたはソフトウェアアプリケーションを実行するように適合されてもよい。サービスまたはソフトウェアアプリケーションは非仮想環境および仮想環境を含み得る。仮想環境は、２次元または３次元（three-dimensional：３Ｄ）表現、ページベースの論理的環境などであろうとなかろうと、仮想イベント、トレードショー、シミュレータ、クラスルーム、購買商品取引および企業活動のために用いられ

るものを含み得る。いくつかの実施形態においては、これらのサービスは、ウェブベースのサービスもしくはクラウドサービスとして、またはソフトウェア・アズ・ア・サービス（Software as a Service：ＳａａＳ）モデルのもとで、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および／または１３０８のユーザに供給されてもよい。そして、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および／または１３０８を動作させるユーザは、１つ以上のクライアントアプリケーションを利用して、サーバ１３１２と相互作用して、これらの構成要素によって提供されるサービスを利用し得る。

【００９８】

図１３に示されている構成では、システム１３００のソフトウェアコンポーネント１３１８，１３２０および１３２２は、サーバ１３１２上に実装されるように示されている。また、他の実施形態においては、システム１３００の構成要素のうちの１つ以上および／またはこれらの構成要素によって提供されるサービスは、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および／または１３０８のうちの１つ以上によって実現されてもよい。その場合、クライアントコンピューティングデバイスを動作させるユーザは、１つ以上のクライアントアプリケーションを利用して、これらの構成要素によって提供されるサービスを使用し得る。これらの構成要素は、ハードウェア、ファームウェア、ソフトウェア、またはそれらの組合せで実現されてもよい。分散型システム１３００とは異なり得るさまざまな異なるシステム構成が可能であることが理解されるべきである。したがって、図１３に示されている実施形態は、実施形態のシステムを実現するための分散型システムの一例であり、限定的であるよう意図されたものではない。

【００９９】

クライアントコンピューティングデバイス１３０２，１３０４，１３０６および／または１３０８は、手持ち式携帯機器（たとえばｉＰｈｏｎｅ（登録商標）、携帯電話、ｉＰａｄ（登録商標）、計算タブレット、パーソナルデジタルアシスタント（personal digital assistant：ＰＤＡ））またはウェアラブル装置（たとえばグーグルグラス（登録商標）ヘッドマウントディスプレイ）であってもよく、当該装置は、マイクロソフトウィンドウズ（登録商標）・モバイル（登録商標）などのソフトウェアを実行し、および／または、ｉＯＳ、ウィンドウズ・フォン、アンドロイド、ブラックベリー１０、パームＯＳなどのさまざまなモバイルオペレーティングシステムを実行し、インターネット、ｅメール、ショート・メッセージ・サービス（short message service：ＳＭＳ）、ブラックベリー（登録商標）、または使用可能な他の通信プロトコルである。クライアントコンピューティングデバイスは、汎用パーソナルコンピュータであってもよく、当該汎用パーソナルコンピュータは、一例として、マイクロソフトウィンドウズ（登録商標）、アップルマッキントッシュ（登録商標）および／またはリナックス（登録商標）オペレーティングシステムのさまざまなバージョンを実行するパーソナルコンピュータおよび／またはラップトップコンピュータを含む。クライアントコンピューティングデバイスは、ワークステーションコンピュータであってもよく、当該ワークステーションコンピュータは、たとえばＧｏｏｇｌｅ　Ｃｈｒｏｍｅ　ＯＳなどのさまざまなＧＮＵ／リナックスオペレーティングシステムを含むがこれらに限定されるものではないさまざまな市販のＵＮＩＸ（登録商標）またはＵＮＩＸライクオペレーティングシステムのうちのいずれかを実行する。代替的には、または付加的には、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および１３０８は、シン・クライアントコンピュータ、インターネットにより可能なゲームシステム（たとえばキネクト（登録商標）ジェスチャ入力装置を備えるかまたは備えないマイクロソフトＸボックスゲーム機）、および／または、ネットワーク１３１０を介して通信が可能なパーソナルメッセージング装置などのその他の電子装置であってもよい。

【０１００】

例示的な分散型システム１３００は、４個のクライアントコンピューティングデバイスを有するように示されているが、任意の数のクライアントコンピューティングデバイスがサポートされてもよい。センサを有する装置などの他の装置が、サーバ１３１２と相互作用

してもよい。
【０１０１】
分散型システム１３００におけるネットワーク１３１０は、さまざまな市販のプロトコルのうちのいずれかを用いてデータ通信をサポートすることができる、当業者になじみのある任意のタイプのネットワークであってもよく、当該プロトコルは、ＴＣＰ／ＩＰ（伝送制御プロトコル／インターネットプロトコル）、ＳＮＡ（システムネットワークアーキテクチャ）、ＩＰＸ（インターネットパケット交換）、アップルトークなどを含むが、これらに限定されるものではない。単に一例として、ネットワーク１３１０は、イーサネット（登録商標）、トークンリングなどに基づくものなどのローカルエリアネットワーク（ＬＡＮ）であってもよい。ネットワーク１３１０は、広域ネットワークおよびインターネットであってもよい。ネットワーク１３１０は、仮想ネットワークを含んでいてもよく、当該仮想ネットワークは、仮想プライベートネットワーク（virtual private network：ＶＰＮ）、イントラネット、エクストラネット、公衆交換電話網（public switched telephone network：ＰＳＴＮ）、赤外線ネットワーク、無線ネットワーク（たとえば米国電気電子学会（Institute of Electrical and Electronics：ＩＥＥＥ）８０２．１１の一連のプロトコル、ブルートゥース（登録商標）および／またはその他の無線プロトコルのうちのいずれかのもとで動作するネットワーク）、および／またはこれらの任意の組合せ、および／または他のネットワークを含むが、これらに限定されるものではない。
【０１０２】
サーバ１３１２は、１つ以上の汎用コンピュータ、専用サーバコンピュータ（一例として、ＰＣ（パーソナルコンピュータ）サーバ、ＵＮＩＸ（登録商標）サーバ、ミッドレンジサーバ、メインフレームコンピュータ、ラックマウント式サーバなどを含む）、サーバファーム、サーバクラスタ、またはその他の適切な構成および／または組合せで構成され得る。サーバ１３１２は、仮想オペレーティングシステムを実行する１つ以上の仮想マシン、または仮想化を含む他のコンピューティングアーキテクチャを含み得る。論理記憶装置の１つ以上のフレキシブルプールは、サーバのための仮想記憶デバイスを維持するように仮想化することができる。仮想ネットワークは、ソフトウェア定義型ネットワーキングを用いて、サーバ１３１２によって制御することができる。さまざまな実施形態においては、サーバ１３１２は、上記の開示に記載されている１つ以上のサービスまたはソフトウェアアプリケーションを実行するように適合され得る。たとえば、サーバ１３１２は、本開示の実施形態に係る上記の処理を実行するためのサーバに対応してもよい。
【０１０３】
サーバ１３１２は、上記のもののうちのいずれか、および、任意の市販のサーバオペレーティングシステムを含むオペレーティングシステムを実行し得る。また、サーバ１３１２は、ＨＴＴＰ（ハイパーテキスト転送プロトコル）サーバ、ＦＴＰ（ファイル転送プロトコル）サーバ、ＣＧＩ（共通ゲートウェイインターフェース）サーバ、ＪＡＶＡ（登録商標）サーバ、データベースサーバなどを含むさまざまな付加的サーバアプリケーションおよび／または中間層アプリケーションのうちのいずれかを実行し得る。例示的なデータベースサーバは、オラクル社（Oracle）、マイクロソフト社（Microsoft）、サイベース社（Sybase）、ＩＢＭ社（International Business Machines）などから市販されているものを含むが、これらに限定されるものではない。
【０１０４】
いくつかの実現例では、サーバ１３１２は、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および１３０８のユーザから受信されたデータフィードおよび／またはイベント更新を分析および統合するための１つ以上のアプリケーションを含み得る。一例として、データフィードおよび／またはイベント更新は、１つ以上の第三者情報源および連続的なデータストリームから受信されるツイッター（登録商標）フィード、フェースブック（登録商標）更新またはリアルタイム更新を含み得るが、これらに限定されるものではなく、センサデータアプリケーション、金融ティッカ、ネットワーク性能測定ツール（たとえばネットワークモニタリングおよびトラフィック管理アプリケーション

）、クリックストリーム分析ツール、自動車交通モニタリングなどに関連するリアルタイムイベントを含み得る。また、サーバ１３１２は、クライアントコンピューティングデバイス１３０２，１３０４，１３０６および１３０８の１つ以上の表示装置を介してデータフィードおよび／またはリアルタイムイベントを表示するための１つ以上のアプリケーションを含み得る。

【０１０５】

また、分散型システム１３００は、１つ以上のデータベース１３１４および１３１６を含み得る。データベース１３１４および１３１６は、さまざまな場所に存在し得る。一例として、データベース１３１４および１３１６の１つ以上は、サーバ１３１２にローカルな（および／または存在する）非一時的な記憶媒体に存在していてもよい。代替的に、データベース１３１４および１３１６は、サーバ１３１２から遠く離れていて、ネットワークベースまたは専用の接続を介してサーバ１３１２と通信してもよい。一組の実施形態においては、データベース１３１４および１３１６は、記憶領域ネットワーク（storage-area network：ＳＡＮ）に存在していてもよい。同様に、サーバ１３１２に起因する機能を実行するための任意の必要なファイルが、サーバ１３１２上にローカルに、および／または、リモートで適宜格納されていてもよい。一組の実施形態においては、データベース１３１４および１３１６は、ＳＱＬフォーマットコマンドに応答してデータを格納、更新および検索するように適合された、オラクル社によって提供されるデータベースなどのリレーショナルデータベースを含み得る。

【０１０６】

図１４は、本開示の実施形態に係る、実施形態のシステムの１つ以上の構成要素によって提供されるサービスをクラウドサービスとして供給することができるシステム環境１４００の１つ以上の構成要素の簡略化されたブロック図である。システム環境１４００は、上述したように、パイプ漏れ予測システムの実施形態を含み得るかまたは実現し得る。示されている実施形態においては、システム環境１４００は、クラウドサービスを提供するクラウドインフラストラクチャシステム１４０２と相互作用するようにユーザによって使用され得る１つ以上のクライアントコンピューティングデバイス１４０４，１４０６および１４０８を含む。クライアントコンピューティングデバイスは、クラウドインフラストラクチャシステム１４０２によって提供されるサービスを使用するためにクラウドインフラストラクチャシステム１４０２と相互作用するようにクライアントコンピューティングデバイスのユーザによって使用され得る、ウェブブラウザ、専有のクライアントアプリケーション（たとえばオラクルフォームズ）または他のアプリケーションなどのクライアントアプリケーションを動作させるように構成され得る。

【０１０７】

図１４に示されているクラウドインフラストラクチャシステム１４０２が図示されている構成要素とは他の構成要素を有し得ることが理解されるべきである。さらに、図１４に示されている実施形態は、本発明の実施形態を組込むことができるクラウドインフラストラクチャシステムの一例に過ぎない。たとえば、クラウドインフラストラクチャシステム１４０２は、上述のように、パイプ漏れ予測システムのうちの１つ以上の要素を含み得るかまたは実現し得る。いくつかの他の実施形態においては、クラウドインフラストラクチャシステム１４０２は、図１４に示されているものよりも多いまたは少ない数の構成要素を有していてもよく、２つ以上の構成要素を組合せてもよく、または構成要素の異なる構成または配置を有していてもよい。

【０１０８】

クライアントコンピューティングデバイス１４０４，１４０６および１４０８は、１３０２，１３０４，１３０６および１３０８について上記したものと類似のデバイスであってもよい。

【０１０９】

例示的なシステム環境１４００は３個のクライアントコンピューティングデバイスを有するように示されているが、任意の数のクライアントコンピューティングデバイスがサポー

トされてもよい。センサなどを有する装置などの他の装置が、クラウドインフラストラクチャシステム１４０２と相互作用してもよい。

【０１１０】

ネットワーク１４１０は、クライアント１４０４，１４０６および１４０８とクラウドインフラストラクチャシステム１４０２との間のデータの通信およびやりとりを容易にし得る。各々のネットワークは、ネットワーク１３１０について上記したものを含むさまざまな市販のプロトコルのうちのいずれかを用いてデータ通信をサポートすることができる、当業者になじみのある任意のタイプのネットワークであってもよい。

【０１１１】

クラウドインフラストラクチャシステム１４０２は、サーバ１３１２について上記したものを含み得る１つ以上のコンピュータおよび／またはサーバを備え得る。

【０１１２】

特定の実施形態においては、クラウドインフラストラクチャシステムによって提供されるサービスは、オンラインデータ記憶およびバックアップソリューション、ウェブベースのｅメールサービス、ホスト型オフィススイートおよびドキュメントコラボレーションサービス、データベース処理、管理技術サポートサービスなどの、クラウドインフラストラクチャシステムのユーザがオンデマンドで利用可能な多数のサービスを含み得る。クラウドインフラストラクチャシステムによって提供されるサービスは、そのユーザのニーズを満たすように動的にスケーリング可能である。クラウドインフラストラクチャシステムによって提供されるサービスの具体的なインスタンス化は、本明細書では「サービスインスタンス」と称される。一般に、インターネットなどの通信ネットワークを介してクラウドサービスプロバイダのシステムからユーザが利用可能な任意のサービスは、「クラウドサービス」と称される。通常、パブリッククラウド環境では、クラウドサービスプロバイダのシステムを構成するサーバおよびシステムは、顧客自身のオンプレミスサーバおよびシステムとは異なっている。たとえば、クラウドサービスプロバイダのシステムがアプリケーションをホストしてもよく、ユーザは、インターネットなどの通信ネットワークを介してオンデマンドで当該アプリケーションを注文および使用してもよい。

【０１１３】

いくつかの例では、コンピュータネットワーククラウドインフラストラクチャにおけるサービスは、ストレージ、ホスト型データベース、ホスト型ウェブサーバ、ソフトウェアアプリケーションへの保護されたコンピュータネットワークアクセス、またはクラウドベンダによってユーザに提供されるかもしくはそうでなければ当該技術分野において公知の他のサービスを含み得る。たとえば、サービスは、インターネットを介したクラウド上のリモートストレージへのパスワードによって保護されたアクセスを含み得る。別の例として、サービスは、ネットワーク化された開発者による私的使用のためのウェブサービスベースのホスト型リレーショナルデータベースおよびスクリプト言語ミドルウェアエンジンを含み得る。別の例として、サービスは、クラウドベンダのウェブサイト上でホストされるｅメールソフトウェアアプリケーションへのアクセスを含み得る。

【０１１４】

特定の実施形態においては、クラウドインフラストラクチャシステム１４０２は、セルフサービスの、サブスクリプションベースの、弾性的にスケーラブルな、信頼性のある、高可用性の、安全な態様で顧客に配信される一連のアプリケーション、ミドルウェアおよびデータベースサービス提供品を含み得る。このようなクラウドインフラストラクチャシステムの一例は、本譲受人によって提供されるオラクルパブリッククラウドである。

【０１１５】

時としてビッグデータとも称される大量のデータは、インフラストラクチャシステムによって、多数のレベルにおいて、および異なるスケールでホストおよび／または操作され得る。このようなデータが含み得るデータセットは、非常に大型で複雑であるので、典型的なデータベース管理ツールまたは従来のデータ処理アプリケーションを用いて処理するのが困難になる可能性がある。たとえば、テラバイトのデータはパーソナルコンピュータま

たはそれらのラックベースの対応物を用いて格納、検索取得および処理することが難しいかもしれない。このようなサイズのデータは、最新のリレーショナルデータベース管理システムおよびデスクトップ統計ならびに視覚化パッケージを用いて機能させるのが困難である可能性がある。それらは、データを許容可能な経過時間内に捕捉しキュレーションし管理し処理するよう、一般的に用いられるソフトウェアツールの構造を超えて、何千ものサーバコンピュータを動作させる大規模並列処理ソフトウェアを必要とし得る。

【０１１６】

大量のデータを視覚化し、トレンドを検出し、および／または、データと相互作用させるために、分析者および研究者は極めて大きいデータセットを格納し処理することができる。平行にリンクされた何十、何百または何千ものプロセッサがこのようなデータに対して作用可能であり、これにより、このようなデータを表示し得るか、または、データに対する外力をシミュレートし得るかもしくはそれが表しているものをシミュレートし得る。これらのデータセットは、データベースにおいて編制されたデータ、もしくは構造化モデルに従ったデータ、および／または、非体系的なデータ（たとえば電子メール、画像、データブロブ（バイナリ大型オブジェクト）、ウェブページ、複雑なイベント処理）などの構造化されたデータを必要とする可能性がある。目標物に対してより多くの（またはより少数の）コンピューティングリソースを比較的迅速に集中させるために実施形態の能力を強化することにより、ビジネス、政府関係機関、研究組織、私人、同じ目的をもった個々人もしくは組織のグループ、または他のエンティティからの要求に基づいて大量のデータセット上でタスクを実行するために、クラウドインフラストラクチャシステムがより良好に利用可能となる。

【０１１７】

さまざまな実施形態においては、クラウドインフラストラクチャシステム１４０２は、クラウドインフラストラクチャシステム１４０２によって供給されるサービスへの顧客のサブスクリプションを自動的にプロビジョニング、管理および追跡するように適合され得る。クラウドインフラストラクチャシステム１４０２は、さまざまなデプロイメントモデルを介してクラウドサービスを提供し得る。たとえば、クラウドインフラストラクチャシステム１４０２が、（たとえばオラクル社によって所有される）クラウドサービスを販売する組織によって所有され、一般大衆またはさまざまな産業企業がサービスを利用できるパブリッククラウドモデルのもとでサービスが提供されてもよい。別の例として、クラウドインフラストラクチャシステム１４０２が単一の組織のためだけに運営され、当該組織内の１つ以上のエンティティにサービスを提供し得るプライベートクラウドモデルのもとでサービスが提供されてもよい。また、クラウドインフラストラクチャシステム１４０２およびクラウドインフラストラクチャシステム１４０２によって提供されるサービスが、関連のコミュニティ内のいくつかの組織によって共有されるコミュニティクラウドモデルのもとでクラウドサービスが提供されてもよい。また、２つ以上の異なるモデルの組合せであるハイブリッドクラウドモデルのもとでクラウドサービスが提供されてもよい。

【０１１８】

いくつかの実施形態においては、クラウドインフラストラクチャシステム１４０２によって提供されるサービスは、ソフトウェア・アズ・ア・サービス（Software as a Service：ＳａａＳ）カテゴリ、プラットフォーム・アズ・ア・サービス（Platform as a Service：ＰａａＳ）カテゴリ、インフラストラクチャ・アズ・ア・サービス（Infrastructure as a Service：ＩａａＳ）カテゴリ、またはハイブリッドサービスを含むサービスの他のカテゴリのもとで提供される１つ以上のサービスを含み得る。顧客は、サブスクリプションオーダーによって、クラウドインフラストラクチャシステム１４０２によって提供される１つ以上のサービスを注文し得る。次いで、クラウドインフラストラクチャシステム１４０２は、顧客のサブスクリプションオーダーでサービスを提供するために処理を実行する。

【０１１９】

いくつかの実施形態においては、クラウドインフラストラクチャシステム１４０２によっ

て提供されるサービスは、アプリケーションサービス、プラットフォームサービスおよびインフラストラクチャサービスを含み得るが、これらに限定されるものではない。いくつかの例では、アプリケーションサービスは、ＳａａＳプラットフォームを介してクラウドインフラストラクチャシステムによって提供されてもよい。ＳａａＳプラットフォームは、ＳａａＳカテゴリに分類されるクラウドサービスを提供するように構成され得る。たとえば、ＳａａＳプラットフォームは、一体化された開発およびデプロイメントプラットフォーム上で一連のオンデマンドアプリケーションを構築および配信するための機能を提供し得る。ＳａａＳプラットフォームは、ＳａａＳサービスを提供するための基本的なソフトウェアおよびインフラストラクチャを管理および制御し得る。ＳａａＳプラットフォームによって提供されるサービスを利用することによって、顧客は、クラウドインフラストラクチャシステムで実行されるアプリケーションを利用することができる。顧客は、顧客が別々のライセンスおよびサポートを購入する必要なく、アプリケーションサービスを取得することができる。さまざまな異なるＳａａＳサービスが提供されてもよい。例としては、大規模組織のための販売実績管理、企業統合およびビジネスの柔軟性のためのソリューションを提供するサービスが挙げられるが、これらに限定されるものではない。

【０１２０】

いくつかの実施形態においては、プラットフォームサービスは、ＰａａＳプラットフォームを介してクラウドインフラストラクチャシステムによって提供されてもよい。ＰａａＳプラットフォームは、ＰａａＳカテゴリに分類されるクラウドサービスを提供するように構成され得る。プラットフォームサービスの例としては、組織（オラクル社など）が既存のアプリケーションを共有の共通アーキテクチャ上で統合することを可能にするサービス、および、プラットフォームによって提供される共有のサービスを活用する新たなアプリケーションを構築する機能を挙げることができるが、これらに限定されるものではない。ＰａａＳプラットフォームは、ＰａａＳサービスを提供するための基本的なソフトウェアおよびインフラストラクチャを管理および制御し得る。顧客は、顧客が別々のライセンスおよびサポートを購入する必要なく、クラウドインフラストラクチャシステムによって提供されるＰａａＳサービスを取得することができる。プラットフォームサービスの例としては、オラクルＪａｖａクラウドサービス（Java Cloud Service：ＪＣＳ）、オラクルデータベースクラウドサービス（Database Cloud Service：ＤＢＣＳ）などが挙げられるが、これらに限定されるものではない。

【０１２１】

ＰａａＳプラットフォームによって提供されるサービスを利用することによって、顧客は、クラウドインフラストラクチャシステム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利用することができ、デプロイされたサービスを制御することもできる。いくつかの実施形態においては、クラウドインフラストラクチャシステムによって提供されるプラットフォームサービスは、データベースクラウドサービス、ミドルウェアクラウドサービル（たとえばオラクルフージョンミドルウェアサービス）およびＪａｖａクラウドサービスを含み得る。一実施形態においては、データベースクラウドサービスは、組織がデータベースリソースをプールしてデータベースクラウドの形態でデータベース・アズ・ア・サービスを顧客に供給することを可能にする共有のサービスデプロイメントモデルをサポートし得る。ミドルウェアクラウドサービスは、クラウドインフラストラクチャシステムにおいてさまざまなビジネスアプリケーションを開発およびデプロイするために顧客にプラットフォームを提供し得るともに、Ｊａｖａクラウドサービスは、クラウドインフラストラクチャシステムにおいてＪａｖａアプリケーションをデプロイするために顧客にプラットフォームを提供し得る。

【０１２２】

さまざまな異なるインフラストラクチャサービスは、クラウドインフラストラクチャシステムにおけるＩａａＳプラットフォームによって提供されてもよい。インフラストラクチャサービスは、ストレージ、ネットワークなどの基本的な計算リソース、ならびに、ＳａａＳプラットフォームおよびＰａａＳプラットフォームによって提供されるサービスを利

用する顧客のための他の基礎的な計算リソースの管理および制御を容易にする。
【０１２３】
また、特定の実施形態においては、クラウドインフラストラクチャシステム１４０２は、クラウドインフラストラクチャシステムの顧客にさまざまなサービスを提供するために使用されるリソースを提供するためのインフラストラクチャリソース１４３０を含み得る。一実施形態においては、インフラストラクチャリソース１４３０は、ＰａａＳプラットフォームおよびＳａａＳプラットフォームによって提供されるサービスを実行するための、サーバ、ストレージおよびネットワーキングリソースなどのハードウェアの予め一体化された最適な組合せを含み得る。
【０１２４】
いくつかの実施形態においては、クラウドインフラストラクチャシステム１４０２におけるリソースは、複数のユーザによって共有され、デマンドごとに動的に再割り振りされてもよい。また、リソースは、異なる時間帯にユーザに割り振られてもよい。たとえば、クラウドインフラストラクチャシステム１４３０は、第１の時間帯におけるユーザの第１の組が規定の時間にわたってクラウドインフラストラクチャシステムのリソースを利用することを可能にし得るとともに、異なる時間帯に位置するユーザの別の組への同一のリソースの再割り振りを可能にし得ることによって、リソースの利用を最大化することができる。
【０１２５】
特定の実施形態においては、クラウドインフラストラクチャシステム１４０２のさまざまな構成要素またはモジュール、および、クラウドインフラストラクチャシステム１４０２によって提供されるサービス、によって共有されるいくつかの内部共有サービス１４３２が提供され得る。これらの内部共有サービスは、セキュリティおよびアイデンティティサービス、インテグレーションサービス、企業リポジトリサービス、企業マネージャサービス、ウイルススキャンおよびホワイトリストサービス、高可用性・バックアップおよび回復サービス、クラウドサポートを可能にするためのサービス、ｅメールサービス、通知サービス、ファイル転送サービスなどを含み得るが、これらに限定されるものではない。
【０１２６】
特定の実施形態においては、クラウドインフラストラクチャシステム１４０２は、クラウドインフラストラクチャシステムにおけるクラウドサービス（たとえばＳａａＳサービス、ＰａａＳサービスおよびＩａａＳサービス）の包括的管理を提供し得る。一実施形態においては、クラウド管理機能は、クラウドインフラストラクチャシステム１４０２によって受信された顧客のサブスクリプションをプロビジョニング、管理および追跡などするための機能を含み得る。
【０１２７】
一実施形態においては、図１４に示されるように、クラウド管理機能は、オーダー管理モジュール１４２０、オーダーオーケストレーションモジュール１４２２、オーダープロビジョニングモジュール１４２４、オーダー管理および監視モジュール１４２６、ならびにアイデンティティ管理モジュール１４２８などの１つ以上のモジュールによって提供され得る。これらのモジュールは、汎用コンピュータ、専用サーバコンピュータ、サーバファーム、サーバクラスタ、またはその他の適切な構成および／もしくは組み合わせであり得る１つ以上のコンピュータおよび／またはサーバを含み得るか、またはそれらを用いて提供され得る。
【０１２８】
例示的な動作１４３４において、クライアントデバイス１４０４，１４０６または１４０８などのクライアントデバイスを用いる顧客は、クラウドインフラストラクチャシステム１４０２によって提供される１つ以上のサービスを要求し、クラウドインフラストラクチャシステム１４０２によって供給される１つ以上のサービスのサブスクリプションについてオーダーを行うことによって、クラウドインフラストラクチャシステム１４０２と対話し得る。特定の実施形態においては、顧客は、クラウドユーザインターフェース（User Interface：ＵＩ）、すなわちクラウドＵＩ１４１２、クラウドＵＩ１４１４および／ま

たはクラウドＵＩ１４１６にアクセスして、これらのＵＩを介してサブスクリプションオーダーを行い得る。顧客がオーダーを行ったことに応答してクラウドインフラストラクチャシステム１４０２によって受信されたオーダー情報は、顧客と、顧客がサブスクライブする予定のクラウドインフラストラクチャシステム１４０２によって提供される１つ以上のサービスとを特定する情報を含み得る。

【０１２９】

オーダーが顧客によって行われた後、オーダー情報は、クラウドＵＩ１４１２，１４１４および／または１４１６を介して受信される。

【０１３０】

動作１４３６において、オーダーは、オーダーデータベース１４１８に格納される。オーダーデータベース１４１８は、クラウドインフラストラクチャシステム１４１８によって動作されるとともに他のシステム要素と連携して動作されるいくつかのデータベースのうちの１つであってもよい。

【０１３１】

動作１４３８において、オーダー情報は、オーダー管理モジュール１４２０に転送される。いくつかの例では、オーダー管理モジュール１４２０は、オーダーの確認および確認時のオーダーの予約などのオーダーに関連する請求書発行機能および会計経理機能を実行するように構成され得る。

【０１３２】

動作１４４０において、オーダーに関する情報は、オーダーオーケストレーションモジュール１４２２に通信される。オーダーオーケストレーションモジュール１４２２は、顧客によって行われたオーダーについてのサービスおよびリソースのプロビジョニングをオーケストレートするためにオーダー情報を利用し得る。いくつかの例では、オーダーオーケストレーションモジュール１４２２は、オーダープロビジョニングモジュール１４２４のサービスを用いてサブスクライブされたサービスをサポートするためにリソースのプロビジョニングをオーケストレートし得る。

【０１３３】

特定の実施形態においては、オーダーオーケストレーションモジュール１４２２は、各々のオーダーに関連付けられるビジネスプロセスの管理を可能にし、ビジネス論理を適用してオーダーがプロビジョニングに進むべきか否かを判断する。動作１４４２において、新たなサブスクリプションについてのオーダーを受信すると、オーダーオーケストレーションモジュール１４２２は、リソースを割り振って当該サブスクリプションオーダーを満たすのに必要とされるそれらのリソースを構成するための要求をオーダープロビジョニングモジュール１４２４に送る。オーダープロビジョニングモジュール１４２４は、顧客によってオーダーされたサービスについてのリソースの割り振りを可能にする。オーダープロビジョニングモジュール１４２４は、クラウドインフラストラクチャシステム１４００によって提供されるクラウドサービスと、要求されたサービスを提供するためのリソースをプロビジョニングするために使用される物理的実装層との間にあるレベルの抽象化を提供する。したがって、オーダーオーケストレーションモジュール１４２２は、サービスおよびリソースが実際に実行中にプロビジョニングされるか、事前にプロビジョニングされて要求があったときに割振られる／割当てられるのみであるかなどの実装の詳細から切り離すことができる。

【０１３４】

動作１４４４において、サービスおよびリソースがプロビジョニングされると、提供されたサービスの通知が、クラウドインフラストラクチャシステム１４０２のオーダープロビジョニングモジュール１４２４によってクライアントデバイス１４０４，１４０６および／または１４０８上の顧客に送られ得る。

【０１３５】

動作１４４６において、顧客のサブスクリプションオーダーが、オーダー管理および監視モジュール１４２６によって管理および追跡され得る。いくつかの例では、オーダー管理

および監視モジュール１４２６は、使用される記憶量、転送されるデータ量、ユーザの数、ならびにシステムアップ時間およびシステムダウン時間などのサブスクリプションオーダーにおけるサービスについての使用統計を収集するように構成され得る。

【０１３６】

特定の実施形態においては、クラウドインフラストラクチャシステム１４００は、アイデンティティ管理モジュール１４２８を含み得る。アイデンティティ管理モジュール１４２８は、クラウドインフラストラクチャシステム１４００におけるアクセス管理および認可サービスなどのアイデンティティサービスを提供するように構成され得る。いくつかの実施形態においては、アイデンティティ管理モジュール１４２８は、クラウドインフラストラクチャシステム１４０２によって提供されるサービスを利用したい顧客についての情報を制御し得る。このような情報は、このような顧客のアイデンティティを認証する情報と、それらの顧客がさまざまなシステムリソース（たとえばファイル、ディレクトリ、アプリケーション、通信ポート、メモリセグメントなど）に対してどのアクションを実行することを認可されるかを記載する情報とを含み得る。また、アイデンティティ管理モジュール１４２８は、各々の顧客についての説明的情報、ならびに、どのようにしておよび誰によってこの説明的情報がアクセスおよび変更され得るかについての情報の管理を含み得る。

【０１３７】

図１５は、本発明のさまざまな実施形態を実現することができる例示的なコンピュータシステム１５００を示す。システム１５００は、上記のコンピュータシステムのうちのいずれかを実現するために使用され得る。たとえば、図１に示されるパイプ漏れ予測システムの要素のすべてまたはいくつかは、システム１５００に含まれ得るかまたは実現され得る。図１５に示されているように、コンピュータシステム１５００は、バスサブシステム１５０２を介していくつかの周辺サブシステムと通信する処理ユニット１５０４を含む。これらの周辺サブシステムは、処理加速ユニット１５０６と、Ｉ／Ｏサブシステム１５０８と、記憶サブシステム１５１８と、通信サブシステム１５２４とを含み得る。記憶サブシステム１５１８は、有形のコンピュータ読取可能な記憶媒体１５２２と、システムメモリ１５１０とを含む。

【０１３８】

バスサブシステム１５０２は、コンピュータシステム１５００のさまざまな構成要素およびサブシステムに、意図されたように互いに通信させるための機構を提供する。バスサブシステム１５０２は、単一のバスとして概略的に示されているが、バスサブシステムの代替的な実施形態は、複数のバスを利用してもよい。バスサブシステム１５０２は、メモリバスまたはメモリコントローラ、周辺バス、およびさまざまなバスアーキテクチャのうちのいずれかを使用する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であってもよい。たとえば、このようなアーキテクチャは、ＩＥＥＥ　Ｐ１３８６．１標準に合わせて製造されたメザニンバスとして実現可能な、業界標準アーキテクチャ（Industry Standard Architecture：ＩＳＡ）バス、マイクロチャネルアーキテクチャ（Micro Channel Architecture：ＭＣＡ）バス、拡張ＩＳＡ（Enhanced ISA：ＥＩＳＡ）バス、ビデオ・エレクトロニクス・スタンダーズ・アソシエーション（Video Electronics Standards Association：ＶＥＳＡ）ローカルバスおよび周辺機器相互接続（Peripheral Component Interconnect：ＰＣＩ）バスを含み得る。

【０１３９】

１つ以上の集積回路（たとえば従来のマイクロプロセッサまたはマイクロコントローラ）として実現可能な処理ユニット１５０４は、コンピュータシステム１５００の動作を制御する。処理ユニット１５０４には、１つ以上のプロセッサが含まれ得る。これらのプロセッサは、単一コアまたはマルチコアのプロセッサを含み得る。特定の実施形態においては、処理ユニット１５０４は、各々の処理ユニットに含まれる単一コアまたはマルチコアのプロセッサを有する１つ以上の独立した処理ユニット１５３２および／または１５３４として実現されてもよい。また、他の実施形態においては、処理ユニット１５０４は、２つのデュアルコアプロセッサを単一のチップに組み入れることによって形成されるクアッド

コア処理ユニットとして実現されてもよい。
【０１４０】
さまざまな実施形態においては、処理ユニット１５０４は、プログラムコードに応答してさまざまなプログラムを実行し得るとともに、同時に実行される複数のプログラムまたはプロセスを維持し得る。任意の所与の時点において、実行されるべきプログラムコードのうちのいくつかまたは全ては、プロセッサ１５０４および／または記憶サブシステム１５１８に存在し得る。好適なプログラミングを通じて、プロセッサ１５０４は、上記のさまざまな機能を提供し得る。また、コンピュータシステム１５００は、加えて、デジタル信号プロセッサ（digital signal processor：ＤＳＰ）、特殊用途プロセッサなどを含み得る処理加速ユニット１５０６を含み得る。
【０１４１】
Ｉ／Ｏサブシステム１５０８は、ユーザインターフェイス入力装置と、ユーザインターフェイス出力装置とを含み得る。ユーザインターフェイス入力装置は、キーボード、マウスまたはトラックボールなどのポインティング装置、タッチパッドまたはタッチスクリーンを含んでいてもよく、これらは、音声コマンド認識システム、マイクロホンおよび他のタイプの入力装置とともに、ディスプレイ、スクロールホイール、クリックホイール、ダイアル、ボタン、スイッチ、キーパッド、オーディオ入力装置に組込まれている。ユーザインターフェイス入力装置は、たとえば、ジェスチャおよび話されたコマンドを用いてナチュラルユーザインターフェースを介してユーザがマイクロソフトＸｂｏｘ（登録商標）３６０ゲームコントローラなどの入力装置を制御して入力装置と対話することを可能にするマイクロソフトキネクト（登録商標）モーションセンサなどのモーション検知および／またはジェスチャ認識装置を含み得る。また、ユーザインターフェイス入力装置は、ユーザから眼球運動（たとえば撮影および／またはメニュー選択を行っている間の「まばたき」）を検出して、当該眼球ジェスチャを入力装置への入力として変換するグーグルグラス（登録商標）まばたき検出器などの眼球ジェスチャ認識装置を含み得る。また、ユーザインターフェイス入力装置は、ユーザが音声コマンドを介して音声認識システム（たとえばＳｉｒｉ（登録商標）ナビゲータ）と対話することを可能にする音声認識検知装置を含み得る。
【０１４２】
また、ユーザインターフェイス入力装置は、三次元（３Ｄ）マウス、ジョイスティックまたはポインティングスティック、ゲームパッドおよびグラフィックタブレット、およびスピーカなどのオーディオ／ビジュアル装置、デジタルカメラ、デジタルカムコーダ、携帯型メディアプレーヤ、ウェブカム、画像スキャナ、指紋スキャナ、バーコードリーダ３Ｄスキャナ、３Ｄプリンタ、レーザレンジファインダ、および視線検出装置を含み得るが、これらに限定されるものではない。また、ユーザインターフェイス入力装置は、たとえば、コンピュータ断層撮影、磁気共鳴画像化、位置発光断層撮影、医療用超音波検査装置などの医療用画像化入力装置を含み得る。また、ユーザインターフェイス入力装置は、たとえばＭＩＤＩキーボード、デジタル楽器などのオーディオ入力装置を含み得る。
【０１４３】
ユーザインターフェイス出力装置は、ディスプレイサブシステム、表示灯、またはオーディオ出力装置などの非視覚的ディスプレイなどを含み得る。ディスプレイサブシステムは、陰極線管（cathode ray tube：ＣＲＴ）、液晶ディスプレイ（liquid crystal display：ＬＣＤ）またはプラズマディスプレイを使用するものなどのフラットパネルディスプレイ、投影装置、タッチスクリーンなどであってもよい。一般に、「出力装置」という用語の使用は、コンピュータシステム１５００からの情報をユーザまたは他のコンピュータに出力するための全ての実現可能なタイプの装置および機構を含むよう意図されている。たとえば、ユーザインターフェイス出力装置は、モニタ、プリンタ、スピーカ、ヘッドホン、自動車のナビゲーションシステム、プロッタ、音声出力装置およびモデムなどの、テキスト、グラフィックスおよびオーディオ／ビデオ情報を視覚的に伝えるさまざまな表示装置を含み得るが、これらに限定されるものではない。

【０１４４】
コンピュータシステム１５００は、現在のところシステムメモリ１５１０内に位置しているように示されているソフトウェア要素を備える記憶サブシステム１５１８を備え得る。システムメモリ１５１０は、処理ユニット１５０４上でロード可能および実行可能なプログラム命令と、これらのプログラムの実行中に生成されるデータとを格納し得る。
【０１４５】
コンピュータシステム１５００の構成およびタイプに応じて、システムメモリ１５１０は、揮発性（ランダムアクセスメモリ（random access memory：ＲＡＭ）など）であってもよく、および／または、不揮発性（リードオンリメモリ（read-only memory：ＲＯＭ）、フラッシュメモリなど）であってもよい。ＲＡＭは、典型的には、処理ユニット１５０４が直ちにアクセス可能なデータおよび／またはプログラムモジュール、および／または、処理ユニット１５０４によって現在動作および実行されているデータおよび／またはプログラムモジュールを収容する。いくつかの実現例では、システムメモリ１５１０は、スタティックランダムアクセスメモリ（static random access memory：ＳＲＡＭ）または
ダイナミックランダムアクセスメモリ（dynamic random access memory：ＤＲＡＭ）などの複数の異なるタイプのメモリを含み得る。いくつかの実現例では、始動中などにコンピュータシステム１５００内の要素間で情報を転送することを助ける基本ルーチンを含む基本入力／出力システム（basic input/output system：ＢＩＯＳ）が、典型的にはＲＯＭ
に格納され得る。一例としておよび非限定的に、システムメモリ１５１０は、クライアントアプリケーション、ウェブブラウザ、中間層アプリケーション、リレーショナルデータベース管理システム（relational database management system：ＲＤＢＭＳ）などを含
み得るアプリケーションプログラム１５１２、プログラムデータ１５１４およびオペレーティングシステム１５１６も示す。一例として、オペレーティングシステム１５１６は、マイクロソフトウィンドウズ（登録商標）、アップルマッキントッシュ（登録商標）および／もしくはリナックスオペレーティングシステムのさまざまなバージョン、さまざまな市販のＵＮＩＸ（登録商標）もしくはＵＮＩＸライクオペレーティングシステム（さまざまなＧＮＵ／リナックスオペレーティングシステム、Ｇｏｏｇｌｅ　Ｃｈｒｏｍｅ（登録商標）ＯＳなどを含むが、これらに限定されるものではない）、ならびに／または、ｉＯＳ、ウィンドウズ（登録商標）フォン、アンドロイド（登録商標）ＯＳ、ブラックベリー（登録商標）１０ＯＳおよびパーム（登録商標）ＯＳオペレーティングシステムなどのモバイルオペレーティングシステムを含み得る。
【０１４６】
また、記憶サブシステム１５１８は、いくつかの実施形態の機能を提供する基本的なプログラミングおよびデータ構造を格納するための有形のコンピュータ読取可能な記憶媒体を提供し得る。プロセッサによって実行されたときに上記の機能を提供するソフトウェア（プログラム、コードモジュール、命令）が記憶サブシステム１５１８に格納され得る。これらのソフトウェアモジュールまたは命令は、処理ユニット１５０４によって実行され得る。また、記憶サブシステム１５１８は、本発明に従って使用されるデータを格納するためのリポジトリを提供し得る。
【０１４７】
また、記憶サブシステム１５００は、コンピュータ読取可能な記憶媒体１５２２にさらに接続可能なコンピュータ読取可能な記憶媒体リーダ１５２０を含み得る。ともにおよび任意には、システムメモリ１５１０と組合せて、コンピュータ読取可能な記憶媒体１５２２は、コンピュータ読取可能な情報を一時的および／または永久に収容、格納、送信および検索するための記憶媒体に加えて、リモートの、ローカルの、固定されたおよび／または取外し可能な記憶装置を包括的に表わし得る。
【０１４８】

コードまたはコードの一部を含むコンピュータ読取可能な記憶媒体１５２２は、当該技術分野において公知のまたは使用される任意の適切な媒体を含み得る。当該媒体は、情報の格納および／または送信のための任意の方法または技術において実現される揮発性および不揮発性の、取外し可能および取外し不可能な媒体などであるが、これらに限定されるものではない記憶媒体および通信媒体を含む。これは、ＲＡＭ、ＲＯＭ、電子的消去・プログラム可能ＲＯＭ（electronically erasable programmable ROM：ＥＥＰＲＯＭ）、フラッシュメモリもしくは他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多用途ディスク（digital versatile disk：ＤＶＤ）、または他の光学式記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置もしくは他の磁気記憶装置、または他の有形のコンピュータ読取可能な媒体などの有形の一時的なコンピュータ読取可能な記憶媒体を含み得る。また、これは、データ信号、データ送信などの無形の一時的なコンピュータ読取可能な媒体、または、所望の情報を送信するために使用可能であるとともに計算システム１５００によってアクセス可能である他の任意の媒体を含み得る。

【０１４９】

一例として、コンピュータ読取可能な記憶媒体１５２２は、取外し不可能な不揮発性磁気媒体から読取るまたは当該媒体に書込むハードディスクドライブ、取外し可能な不揮発性磁気ディスクから読取るまたは当該ディスクに書込む磁気ディスクドライブ、ならびに、ＣＤ　ＲＯＭ、ＤＶＤおよびブルーレイ（登録商標）ディスクまたは他の光学式媒体などの取外し可能な不揮発性光学ディスクから読取るまたは当該ディスクに書込む光学式ディスクドライブを含み得る。コンピュータ読取可能な記憶媒体１５２２は、ジップ（登録商標）ドライブ、フラッシュメモリカード、ユニバーサルシリアルバス（universal serial bus：ＵＳＢ）フラッシュドライブ、セキュアデジタル（secure digital：ＳＤ）カード、ＤＶＤディスク、デジタルビデオテープなどを含み得るが、これらに限定されるものではない。また、コンピュータ読取可能な記憶媒体１５２２は、フラッシュメモリベースのＳＳＤ、企業向けフラッシュドライブ、ソリッドステートＲＯＭなどの不揮発性メモリに基づくソリッドステートドライブ（solid-state drive：ＳＳＤ）、ソリッドステートＲＡＭ、ダイナミックＲＡＭ、スタティックＲＡＭなどの揮発性メモリに基づくＳＳＤ、ＤＲＡＭベースのＳＳＤ、磁気抵抗ＲＡＭ（magnetoresistive RAM：ＭＲＡＭ）ＳＳＤ、およびＤＲＡＭとフラッシュメモリベースのＳＳＤとの組合せを使用するハイブリッドＳＳＤを含み得る。ディスクドライブおよびそれらの関連のコンピュータ読取可能な媒体は、コンピュータ読取可能な命令、データ構造、プログラムモジュールおよび他のデータをコンピュータシステム１５００に提供し得る。

【０１５０】

通信サブシステム１５２４は、他のコンピュータシステムおよびネットワークとのインターフェイスを提供する。通信サブシステム１５２４は、他のシステムからデータを受信したり、コンピュータシステム１５００から他のシステムにデータを送信するためのインターフェイスの役割を果たす。たとえば、通信サブシステム１５２４は、コンピュータシステム１５００がインターネットを介して１つ以上の装置に接続することを可能にし得る。いくつかの実施形態においては、通信サブシステム１５２４は、（たとえば３Ｇ、４ＧまたはＥＤＧＥ（enhanced data rates for global evolution）などの携帯電話技術、高度データネットワーク技術を用いて）無線音声および／またはデータネットワークにアクセスするための無線周波数（radio frequency：ＲＦ）トランシーバコンポーネント、ＷｉＦｉ（ＩＥＥＥ１９０２．１１ファミリ標準または他のモバイル通信技術またはそれらの任意の組合せ）、全地球測位システム（global positioning system：ＧＰＳ）レシーバコンポーネント、および／または、他のコンポーネントを含み得る。いくつかの実施形態においては、通信サブシステム１５２４は、無線インターフェイスに加えて、または無線インターフェイスの代わりに、有線ネットワーク接続（例えばイーサネット）を提供し得る。

【０１５１】

また、いくつかの実施形態においては、通信サブシステム１５２４は、コンピュータシス

テム１５００を使用し得る１人以上のユーザを代表して、構造化されたおよび／または構造化されていないデータフィード１５２６、イベントストリーム１５２８、イベント更新１５３０などの形態で入力通信を受信し得る。

【０１５２】

一例として、通信サブシステム１５２４は、ツイッター（登録商標）フィード、フェースブック（登録商標）更新、リッチ・サイト・サマリ（Rich Site Summary：ＲＳＳ）フィードなどのウェブフィードなどのデータフィード１５２６をリアルタイムでソーシャルメディアネットワークおよび／または他の通信サービスのユーザから受信し、および／または、１つ以上の第三者情報源からリアルタイム更新を受信するように構成され得る。

【０１５３】

加えて、通信サブシステム１５２４は、連続的なデータストリームの形態でデータを受信するように構成され得る。当該データは、連続的である場合もあれば本質的に明確な端部をもたない状態で境界がない場合もあるリアルタイムイベントのイベントストリーム１５２８および／またはイベント更新１５３０を含み得る。連続的なデータを生成するアプリケーションの例としては、たとえばセンサデータアプリケーション、金融ティッカ、ネットワーク性能測定ツール（たとえばネットワークモニタリングおよびトラフィック管理アプリケーション）、クリックストリーム分析ツール、自動車交通モニタリングなどを含み得る。

【０１５４】

また、通信サブシステム１５２４は、構造化されたおよび／または構造化されていないデータフィード１５２６、イベントストリーム１５２８、イベント更新１５３０などを、コンピュータシステム１５００に結合された１つ以上のストリーミングデータソースコンピュータと通信し得る１つ以上のデータベースに出力するように構成され得る。

【０１５５】

コンピュータシステム１５００は、手持ち式携帯機器（たとえばｉＰｈｏｎｅ（登録商標）携帯電話、ｉＰａｄ（登録商標）計算タブレット、ＰＤＡ）、ウェアラブル装置（たとえばグーグルグラス（登録商標）ヘッドマウントディスプレイ）、ＰＣ、ワークステーション、メインフレーム、キオスク、サーバラックまたはその他のデータ処理システムを含むさまざまなタイプのうちの１つであってもよい。

【０１５６】

コンピュータおよびネットワークの絶え間なく変化し続ける性質のために、図１５に示されているコンピュータシステム１５００の説明は、特定の例として意図されているに過ぎない。図１５に示されているシステムよりも多くのまたは少ない数の構成要素を有する多くの他の構成が可能である。たとえば、ハードウェア、ファームウェア、（アプレットを含む）ソフトウェア、または組合せにおいて、カスタマイズされたハードウェアが使用されてもよく、および／または、特定の要素が実装されてもよい。さらに、ネットワーク入力／出力装置などの他のコンピューティングデバイスへの接続が利用されてもよい。本明細書中に提供される開示および教示に基づいて、当業者は、さまざまな実施形態を実現するための他の手段および／または方法を理解するであろう。

【０１５７】

上述の明細書では、本発明の局面は、その具体的な実施形態を参照して記載されているが、本発明はこれに限定されるものではないことを当業者は認識するであろう。上述の発明のさまざまな特徴および局面は、個々にまたは一緒に使用されてもよい。さらに、実施形態は、明細書のより広い精神および範囲から逸脱することなく、本明細書に記載されているものを越えたいくつもの環境およびアプリケーションでも利用可能である。したがって、明細書および図面は、限定的ではなく例示的なものとみなされるべきである。

【図面】

【図１】

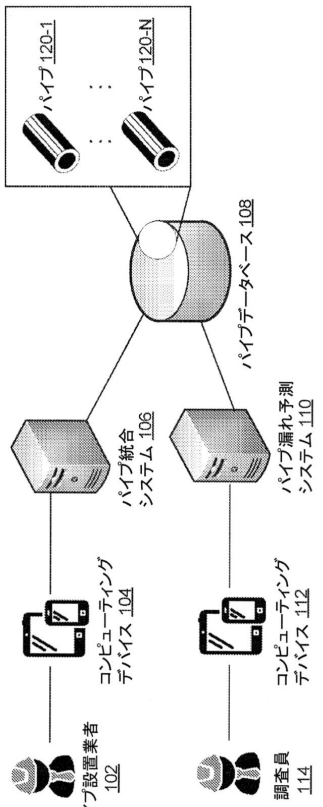

FIG. 1

【図２】

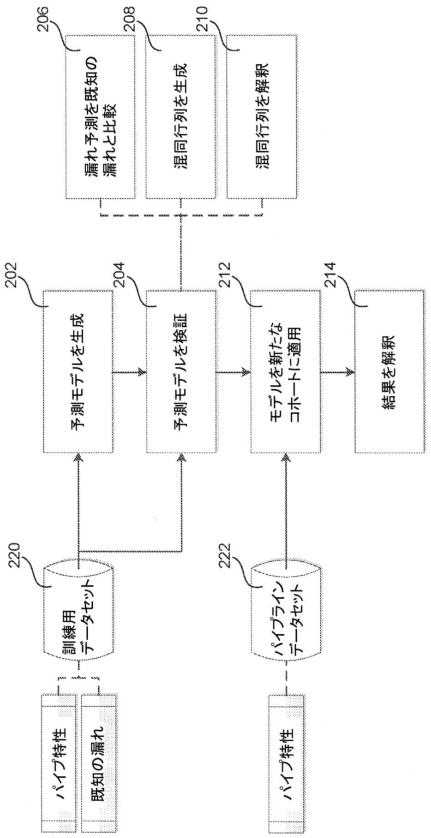

FIG. 2

【図３】

	fid	wsm_nbr	use_type	pipe_diam	len_ft	thickness	material	relined	plr_num	plr_year	soil_res	leak_freq
302		304	306	308	310	312	314	316	318	320	322	324
	9277583	172-147	DIST	12	329.7600	0.71	CI	#NULL#	38754	1934	1138.839	0
	4215359	206-150	DIST	6	664.0825	0.43	CI	CL	71118	1954	3601.720	0
	6253942	124-129	DIST	4	10.8400	0.00	CI	#NULL#	70842	1954	1556.318	0
	9510661	226-144	DIST	20	129.0000	0.31	STL	CL	86973	1968	4668.634	0
	3206401	204-132	DIST	4	440.9578	0.38	CI	CL	76489	1957	1015.702	0
	9148005	170-135	DIST	6	7.4076	0.40	CI	CL	80968	1962	1931.740	0

	leak_yn	district	pipe_age	material_life	pct_life	thickness	pipe_elevation	within_hill
	326	328	330	332	334	336	338	340
1	N	East Valley	82	100	82	333.638	666.362	N
2	N	East Valley	62	100	62	211.726	922.274	N
3	N	Western	62	100	62	275.792	253.208	Y
4	N	East Valley	48	120	40	155.763	1293.24	Y
5	N	West Valley	59	100	59	179.307	954.693	N
6	N	West Valley	54	100	54	317.877	757.123	N

FIG. 3

【図４】

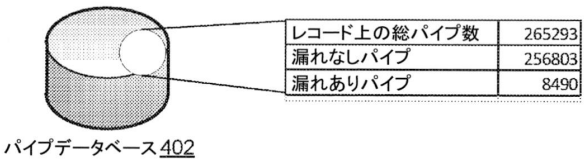

FIG. 4

【図５】

502

予測	参照	
	N	Y
N	256803	107
Y	0	8383

$全体

正答率	Kappa	正答率下側	正答率上側	正答率ヌル	正答率P値	マクネマーP値
9.995967E-01	9.934502E-01	9.995126E-01	9.996695E-01	9.679976E-01	0.000000E+00	1.215611E-24

504

$クラス単位

感度	特異度	陽性適中率	陰性適中率	適合率
0.98739694	1.00000000	1.00000000	0.99958351	1.00000000
再現率	F1	プレバレンス	検出率	検出プレバレンス
0.98739694	0.99365851	0.03200235	0.03159902	0.03159902
平均正答率				
0.99369847				

506 508

FIG. 5

【図６】

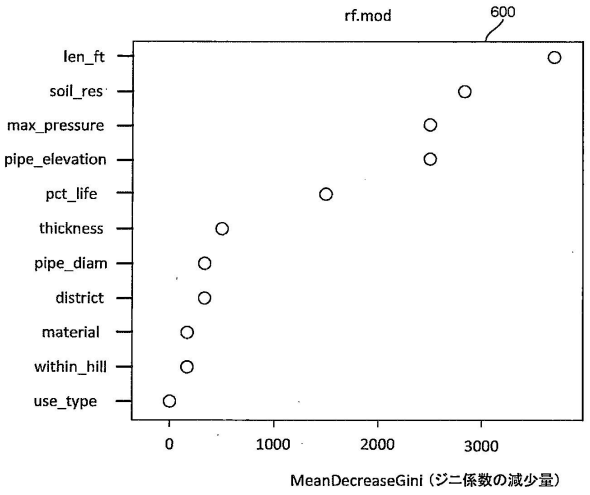

FIG. 6

【図７】

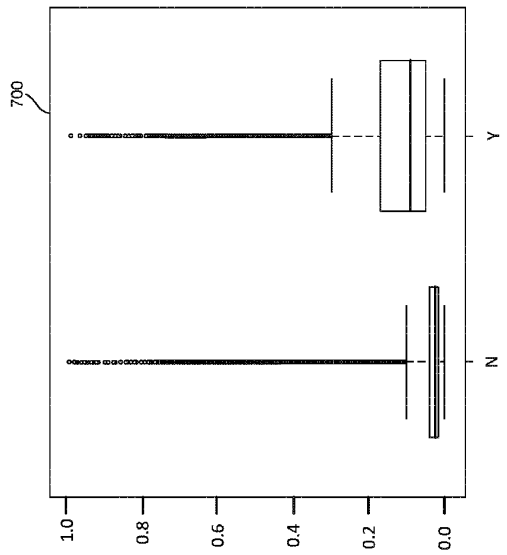

FIG. 7

【図８】

802

予測	参照	
	N	Y
N	196583	2123
Y	60220	6367

$全体

正答率	Kappa	正答率下側	正答率上側	正答率ヌル
0.7650032	0.1196371	0.7633850	0.7666156	0.9679976
正答率P値	マクネマーP値			
1.0000000	0.0000000			

804

808

$クラス単位

感度	特異度	陽性適中率	陰性適中率
0.74994111	0.76550118	0.09561927	0.98931587
適合率	再現率	F1	プレバレンス
0.09561927	0.74994111	0.16961253	0.03200235
検出率	検出プレバレンス	平均正答率	
0.02399988	0.25099418	0.75772114	

806

FIG. 8

【図９】

902

係数:	推定標準誤差	z値	Pr(>|z|)	
（切片）	-1.635E+01	1.582E+02	-0.103	0.917689
use_type#NULL#	1.387E+00	4.620E+02	0.003	0.997604
use_typeAMMONIA	-3.316E+00	6.670E+02	-0.005	0.996033
use_typeCHLOR	-9.752E-01	2.185E+02	-0.004	0.996440
use_typeDECOM	9.977E+00	1.582E+02	0.063	0.949720
use_typeDIST	1.128E+01	1.582E+02	0.071	0.943158
use_typeDRAIN	1.042E+01	1.582E+02	0.066	0.947469
use_typeIRRIG	-1.170E+01	3.346E+02	-0.035	0.972114
use_typeLSLATERAL	1.943E+00	4.382E+02	0.004	0.996463
use_typeMANIFOLD	4.858E-01	1.997E+02	0.002	0.998059
use_typeRECL	1.009E+01	1.582E+02	0.064	0.949156
use_typeTRUNK	1.130E+01	1.582E+02	0.071	0.943064
use_typeWELL	1.244E+01	1.582E+02	0.079	0.937341
pipe_diam	-1.799E-01	5.232E-03	-34.388	< 2e-16 ***
len_ft	3.235E-03	3.684E-05	87.817	< 2e-16 ***
thickness	-7.845E-02	2.140E-02	-3.666	0.000247 ***
materialAC	2.379E-01	1.015E+00	0.234	0.814728
materialBRS	-1.192E+01	5.918E+02	-0.020	0.983933
materialCI	1.248E+00	1.014E+00	1.231	0.218454
materialCM	1.438E+00	1.475E+00	0.975	0.329427
materialCONC	5.840E-01	1.418E+00	0.412	0.680490
materialCOP	2.146E-01	1.021E+00	0.210	0.833561
materialDI	-4.200E-01	1.016E+00	-0.413	0.679330
materialHDPE	3.297E+00	1.331E+00	2.478	0.013229 *
materialPP	4.033E+00	1.061E+03	0.004	0.996967
materialPVC	3.140E+00	1.541E+00	2.038	0.041589 *
materialSTL	5.906E-01	1.015E+00	0.582	0.560635
soil_res	-3.360E-04	1.225E-05	-27.437	< 2e-16 ***
districtCentral	1.339E+00	3.553E-01	3.767	0.000165 ***
districtEast Valley	9.090E-01	3.586E-01	2.535	0.011254 *
districtHarbor	1.166E+00	3.566E-01	3.270	0.001077 **
districtWest Valley	3.182E-01	3.580E-01	0.889	0.374179
districtWestern	1.150E+00	3.556E-01	4.223	2.41e-05 ***
pct_life	1.106E-03	1.470E-04	7.519	5.50e-14 ***
max_pressure	2.442E-03	1.432E-04	17.053	< 2e-16 ***
pipe_elevation	-2.224E-04	6.287E-05	-3.537	0.000405 ***
within_hillY	1.702E-01	2.917E-02	5.836	5.36e-09 ***
重要度コード:	0 *** 0.001 ** 0.01 *			

FIG. 9

【図１０】

1002

予測	参照 N	参照 Y
N	244995	5731
Y	11808	2759

1004

$全体				
正答率	Kappa	正答率下側	正答率上側	正答率スル
0.9338882	0.2072641	0.9329360	0.9348309	0.9679976
正答率P値	マクネマーP値			
1.0000000	0.0000000			

1006

1008

$クラス単位			
感度	特異度	陽性適中率	陰性適中率
0.32497055	0.95401923	0.18940070	0.97714238
適合率	再現率	F1	プレバレンス
0.18940070	0.32496055	0.23931995	0.03200235
検出率	検出プレバレンス	平均正答率	
0.01039982	0.05490910	0.63949489	

FIG. 10

【図１１】

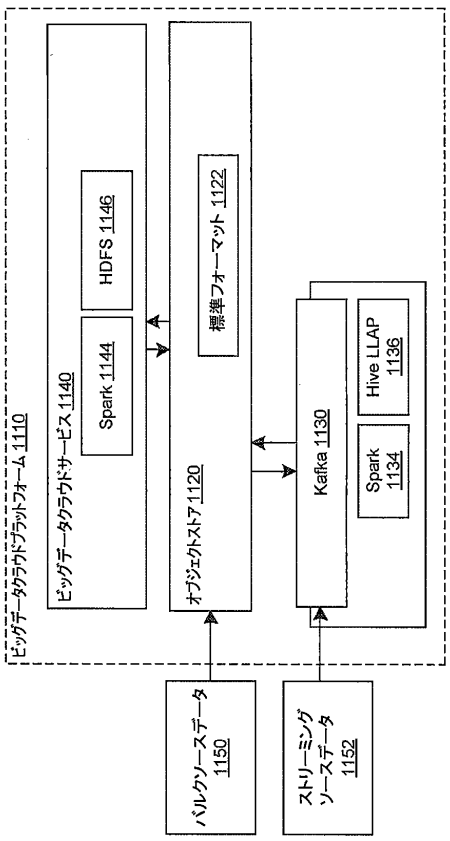

FIG. 11

【図１２】

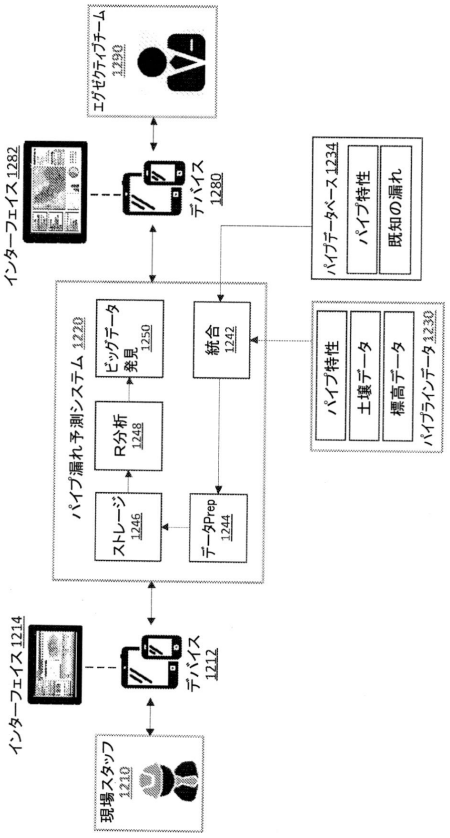

FIG. 12

【図１３】

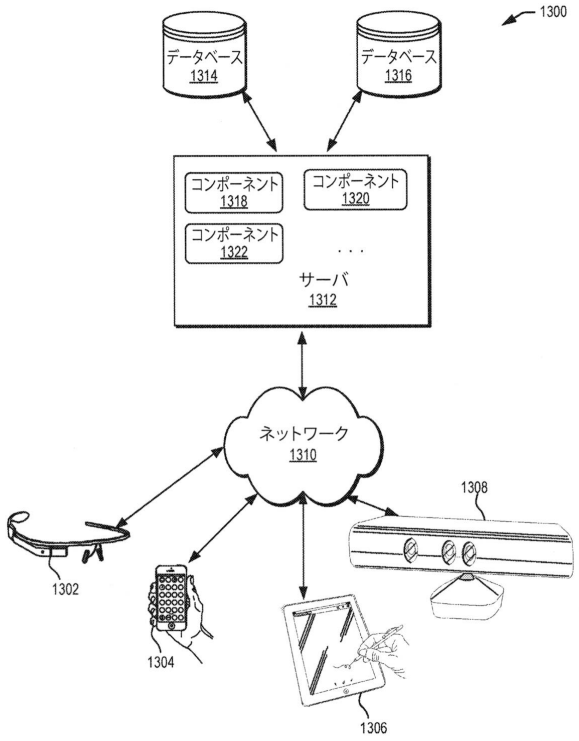

FIG. 13

【図１４】

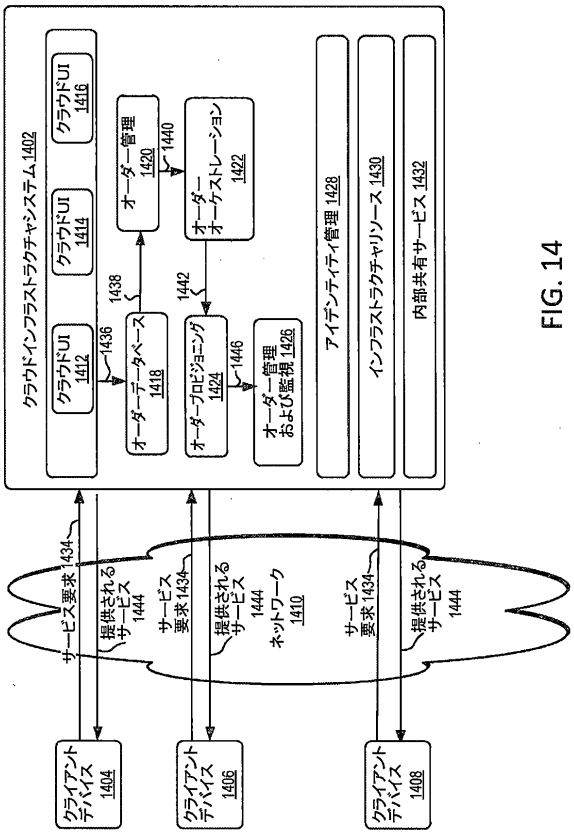

FIG. 14

【図１５】

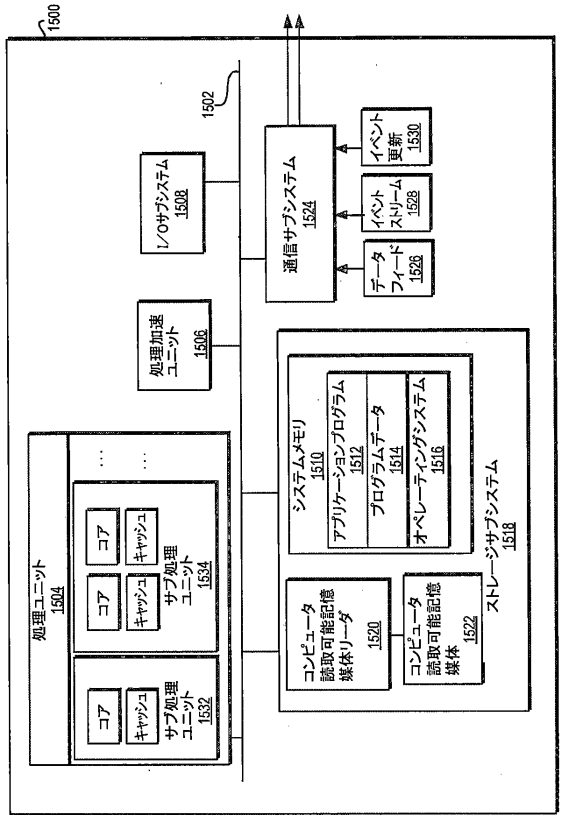

FIG. 15

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)

(56)参考文献
特開２０１２－０１４６１９（ＪＰ，Ａ）
特開２０１１－１９８２６７（ＪＰ，Ａ）
Mohammad Burhan Abdulla et al.，Pipeline leak detection using artificial neuralnetwork: Experimental study，2013 5th InternationalConference on Modelling, Identification and Control (ICMIC)，2013年09月02日，pp.328-332
R.B. Santos et al.，Real-Time Monitoring of Gas Pipeline through Artificial Neural Networks，2013 BRICS Congress on ComputationalIntelligence and 11th Brazilian Congress on Computational Intelligence，2013年，pp.329-334
Mohammed S. El Abbasy et al.，Artificial neural network models for predicting condition of offshore oil and gas pipelines，Automation in Construction 45，2014年，pp.50-65
Rui Wang et al.，Pipe Failure Prediction: A Data Mining Method，ICDE Conference 2013，2013年，pp.1208-1218
Brett Lantz，Rによる機械学習，株式会社翔泳社，2017年03月02日，pp.18-21,284-304,311-313,326-334
辺 松 外３名，機械学習による経年劣化タイミング解析手法，情報処理学会 シンポジウム ＤＡシンポジウム ２０１６，2016年09月07日，pp.44-49
叶賀 卓 外１名，単極脳波信号を用いたトランプゲーム中の子どもの喜怒哀楽予測モデル構築，電気学会研究会資料，2014年09月01日，pp.117-120
下田 倫大，詳解 Ａｐａｃｈｅ Ｓｐａｒｋ 初版，株式会社技術評論社，2016年06月01日，pp.206-211,223

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｑ １０／００ － ９９／００
Ｇ０６Ｎ ２０／００