

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7353238号
(P7353238)

(45)発行日 令和5年9月29日(2023.9.29)

(24)登録日 令和5年9月21日(2023.9.21)

(51)国際特許分類		F I	
G 0 6 N	7/01 (2023.01)	G 0 6 N	7/01
G 0 5 B	19/418 (2006.01)	G 0 5 B	19/418 Z
G 0 5 B	23/02 (2006.01)	G 0 5 B	23/02 3 0 2 R
G 0 6 F	11/07 (2006.01)	G 0 5 B	23/02 3 0 2 Y
		G 0 6 F	11/07 1 4 0 V
請求項の数 17 (全28頁) 最終頁に続く			
(21)出願番号	特願2020-95164(P2020-95164)	(73)特許権者	504407000
(22)出願日	令和2年6月1日(2020.6.1)		パロ アルト リサーチ センター インコ
(65)公開番号	特開2021-2335(P2021-2335A)		ーポレイテッド
(43)公開日	令和3年1月7日(2021.1.7)		アメリカ合衆国 カリフォルニア州 9 4
審査請求日	令和5年5月29日(2023.5.29)		3 0 4 パロ アルト カイオーテ ヒル
(31)優先権主張番号	16/448,273		ロード 3 3 3 3
(32)優先日	令和1年6月21日(2019.6.21)	(74)代理人	100094569
(33)優先権主張国・地域又は機関	米国(US)		弁理士 田中 伸一郎
早期審査対象出願		(74)代理人	100109070
			弁理士 須田 洋之
		(74)代理人	100067013
			弁理士 大塚 文昭
		(74)代理人	100086771
			弁理士 西島 孝喜
		(74)代理人	100109335
		最終頁に続く	

(54)【発明の名称】 高次元センサデータにおける異常事象の自動化された根本原因分析を実行する方法及びシステム

(57)【特許請求の範囲】

【請求項1】

1つ以上の機械の動作における異常の根本原因を識別する方法であって、
前記1つ以上の機械に関連付けられたセンサのセットからセンサデータを取得すること
と、
前記センサデータをセンサ状態のセットに変換することと、を含み、
前記方法は、さらに、
入力のセット及び前記センサ状態のセットに基づいて有向非循環グラフ（DAG）のセ
ットを構築し、
ベイジアンネットワーク発見技術を使用することにより、前記DAGのセットから前記
最適なDAGを検索することによって、
前記センサ状態のセットに基づいて、前記センサのセットに対する最適な前記DAGを
構築して因果関係をモデル化することと、
前記DAGを使用することによって、直接隣接センサの状態に基づいて標的センサの異
常状態の確率を判定することと、
前記DAGにおける前記異常状態をバックトラッキングすることによって、前記標的セ
ンサに関連付けられた前記異常状態の根本原因を判定することと、を備える、方法。
【請求項2】
前記センサデータをセンサ状態のセットに変換することが、
前記センサデータにデータ前処理技術を適用して、前記センサデータを統一されたグロ

ーバル基準時間と時間整合させることであって、前記統一されたグローバル基準時間が時間間隔のセットを含むことと、

前記前処理されたセンサデータにデータ要約技術を適用することと、を備える、請求項 1 に記載の方法。

【請求項 3】

前記データ前処理技術を適用することが、更に、

前記時間整合されたセンサデータにデータ補間技術を適用して、欠落したセンサデータサンプルを置き換えることを備える、請求項 2 に記載の方法。

【請求項 4】

前記データ要約技術を前記前処理されたセンサデータに適用することが、

データ変換技術を適用して、前記統一されたグローバル基準時間に関連付けられた各時間間隔に対するメトリックのセットを生成することと、

前記メトリックのセットに基づいて、クラスタリング技術を使用することによって前記センサのセットに関連付けられた属性のセットにおいて、強力的に接続されたコンポーネントである類似の属性をグループ化し、前記強力的に接続されたコンポーネントのそれぞれを単一のメタノードに縮小させることと、

メタノードのセットに基づいて、前記メタノードの DAG としてメタグラフを構築することと、を備える、請求項 2 に記載の方法。

【請求項 5】

前記入力のセットが、

気象データと、

ユーザ指定のブラックリストと、

ユーザ指定のホワイトリストと、のうちの 1 つ以上を含む、請求項 1 に記載の方法。

【請求項 6】

前記センサ状態のセットが、K ミーンズアルゴリズムを使用することによって取得される、請求項 1 に記載の方法。

【請求項 7】

前記センサのセットの前記 DAG が、

ノードのセットであって、前記グラフ内の各ノードがセンサクラスタのセット内のセンサを表す、ノードのセットと、

有向エッジのセットであって、前記エッジのセット内の各有向エッジが前記グラフ内の 2 つのノード間の因果関係を表す、有向エッジのセットと、を含む、請求項 1 に記載の方法。

【請求項 8】

1 つ以上の機械の動作における異常の根本原因を識別する装置であって、

1 つ以上のプロセッサと、

前記 1 つ以上の機械に埋め込まれたセンサのセットと、

前記 1 つ以上のプロセッサによって実行されると、前記装置に、

前記 1 つ以上の機械に関連付けられたセンサのセットからセンサデータを取得させ、

前記センサデータをセンサ状態のセットに変換させ、

さらに、前記装置に、

入力のセット及び前記センサ状態のセットに基づいて有向非循環グラフ (DAG) のセットを構築し、

ベイジアンネットワーク発見技術を使用することにより、前記 DAG のセットから前記最適な DAG を検索することにより、

前記センサ状態のセットに基づいて、前記センサのセットに対する最適な前記 DAG を構築させて因果関係をモデル化させ、

前記 DAG を使用することによって、直接隣接センサの状態に基づいて標的センサの異常状態の確率を判定させ、

前記 DAG における前記異常状態をバックトラッキングすることによって、前記標的

10

20

30

40

50

センサに関連付けられた前記異常状態の根本原因を判定させる、命令を記憶するメモリと、を備える、装置。

【請求項 9】

前記センサデータをセンサ状態のセットに変換することが、

前記センサデータにデータ前処理技術を適用して、前記センサデータを統一されたグローバル基準時間と時間整合させることであって、前記統一されたグローバル基準時間が時間間隔のセットを含むことと、

前記前処理されたセンサデータにデータ要約技術を適用することと、を備える、請求項 8 に記載の装置。

【請求項 10】

前記データ前処理技術を適用することが、更に、

前記時間整合されたセンサデータにデータ補間技術を適用して、欠落したセンサデータサンプルを置き換えることを備える、請求項 9 に記載の装置。

【請求項 11】

前記データ要約技術を前記前処理されたセンサデータに適用することが、

データ変換技術を適用して、前記統一されたグローバル基準時間に関連付けられた各時間間隔に対するメトリックのセットを生成することと、

前記メトリックのセットに基づいて、クラスタリング技術を使用することによって前記センサのセットに関連付けられた属性のセットにおいて、強力に接続されたコンポーネントである類似の属性をグループ化し、前記強力に接続されたコンポーネントのそれぞれを単一のメタノードに縮小させることと、

メタノードのセットに基づいて、前記メタノードの DAG としてメタグラフを構築することと、を備える、請求項 9 に記載の装置。

【請求項 12】

前記入力のセットが、

気象データと、

ユーザ指定のブラックリストと、

ユーザ指定のホワイトリストと、のうちの 1 つ以上を含む、請求項 8 に記載の装置。

【請求項 13】

前記センサ状態のセットが、K ミーンズアルゴリズムを使用することによって取得される、請求項 8 に記載の装置。

【請求項 14】

前記センサのセットの前記 DAG が、

ノードのセットであって、前記グラフ内の各ノードがセンサクラスタのセット内のセンサを表す、ノードのセットと、

有向エッジのセットであって、前記エッジのセット内の各有向エッジが前記グラフ内の 2 つのノード間の因果関係を表す、有向エッジのセットと、を含む、請求項 8 に記載の装置。

【請求項 15】

コンピュータによって実行されると、前記コンピュータに、1 つ以上の機械の動作における異常の根本原因を識別する方法を実行させる命令を記憶する非一時的コンピュータ可読記憶媒体であって、前記方法が、

前記 1 つ以上の機械に関連付けられたセンサのセットからセンサデータを取得することと、

前記センサデータをセンサ状態のセットに変換することと、を含み、

前記方法は、さらに、

入力のセット及び前記センサ状態のセットに基づいて有向非循環グラフ（DAG）のセットを構築し、

ベイジアンネットワーク発見技術を使用することにより、前記 DAG のセットから前記最適な DAG を検索することによって、

10

20

30

40

50

前記センサ状態のセットに基づいて、前記センサのセットに対する最適な前記DAGを構築して因果関係をモデル化することと、

前記DAGを使用することによって、直接隣接センサの状態に基づいて標的センサの異常状態の確率を判定することと、

前記DAGにおける前記異常状態をバックトラッキングすることによって、前記標的センサに関連付けられた前記異常状態の根本原因を判定することと、を備える、非一時的コンピュータ可読記憶媒体。

【請求項16】

前記センサデータをセンサ状態のセットに変換することが、

前記センサデータにデータ前処理技術を適用して、前記センサデータを統一されたグローバル基準時間と時間整合させることであって、前記統一されたグローバル基準時間が時間間隔のセットを含むことと、

前記前処理されたセンサデータにデータ要約技術を適用することと、を備える、請求項15に記載の非一時的コンピュータ可読記憶媒体。

【請求項17】

前記データ要約技術を前記前処理されたセンサデータに適用することが、

データ変換技術を適用して、前記統一されたグローバル基準時間に関連付けられた各時間間隔に対するメトリックのセットを生成することと、

前記メトリックのセットに基づいて、クラスタリング技術を使用することによって前記センサのセットに関連付けられた属性のセットにおいて、強力に接続されたコンポーネントである類似の属性をグループ化し、前記強力に接続されたコンポーネントのそれぞれを単一のメタノードに縮小させることと、

メタノードのセットに基づいて、前記メタノードのDAGとしてメタグラフを構築することと、を備える、請求項16に記載の非一時的コンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、一般に、異常事象の根本原因分析を実行するシステム及び方法に関する。より具体的には、本開示は、高次元センサデータから隠れた異常事象の根本原因を識別することに関する。

【背景技術】

【0002】

多数の産業用途、例えば、デジタル製造において、モノのインターネット（IoT）対応機器の使用が増加するのに伴い、大量のセンサデータが容易に利用可能である。多くのサイバー物理システム（CPS）用途では、センサデータは、IoT装備装置から連続的に生成される。そのようなセンサデータは、工場機械の動作効率を最適化するために使用されることができる実施可能な情報を含む。具体的には、CPSにおける動作効率の最適化及び障害事象のリスクの低減は、工場のダウンタイムを低減し、製造プロセスにおける生産性を改善するために、それらの直近の用途に対する最も追求された目的の1つであった。しかしながら、センサデータの次元性及びサイズが増加すると、工場レイアウトでの機械の動作中に異常事象の根本原因を識別するために、センサ間の因果関係を手動で試験して識別するために、非常に高価なタスクにすぐになる。

【0003】

実世界用途に十分にスケーリングするためには、センサデータからのセンサ間の因果関係を発見するために、根本原因分析ワークフローを自動化する必要がある。

【発明の概要】

【0004】

本発明の一実施形態によれば、1つ以上の機械の動作における異常の根本原因を識別するシステムが提供される。動作中、システムは、機械に関連付けられたセンサのセットからセンサデータを取得し、センサデータをセンサのセット内の各センサのセンサ状態のセ

10

20

30

40

50

ットに変換し、センサ状態のセットに基づいて、センサのセット間で最適な有向非循環グラフ（DAG）を構築して因果関係をモデル化することができ、直接隣接センサの状態を与えられたユーザ定義型標的センサの異常状態の確率を判定し、グラフ構造における異常状態をバックトラッキングすることによって標的センサに関連付けられた異常状態の根本原因を判定する。

【0005】

この実施形態の変形例では、センサのセット内の各センサのセンサ状態のセットにセンサデータを変換することは、データ前処理技術をセンサデータに適用して、センサデータを、グローバル基準時間と時間整合させることであって、統一されたグローバル基準時間が時間間隔のセットを含むことと、データ要約技術を前処理されたセンサデータに適用することとを備える。

10

【0006】

この実施形態の更なる変形例では、データ前処理技術を適用することは、更に、データ補間技術を時間整合されたセンサデータに適用して、欠落したセンサデータサンプルを置き換えることを備える。

【0007】

この実施形態の更なる変形例では、データ要約技術を前処理されたセンサデータに適用することは、データ変換技術を適用して、統一されたグローバル基準時間に関連付けられた各時間間隔に対するメトリックのセットを生成することと、メトリックのセットに基づいて、クラスタリング技術を使用することによってセンサのセットに関連付けられた属性のセットにおいて、強力的に接続されたコンポーネントである類似の属性をグループ化し、強力的に接続されたコンポーネントのそれぞれを単一のメタノードに縮小させることと、メタノードのセットに基づいて、メタノードのDAGとしてメタグラフを構築することとを備える。有向グラフの強力的に接続されたコンポーネントは、強力的に接続されたサブグラフである。サブグラフ内の各ノードがサブグラフ内の全ての他のノードから到達可能である場合、有向グラフ内のサブグラフは、強力的に接続されることができる。類似の属性のグループは、有向グラフ内の強力的に接続されたコンポーネントとして表すことができる。

20

【0008】

この実施形態の変形例では、センサ状態のセットに基づいて、メタノードのセットの中で最適なDAG構造を構築することは、入力セット及びセンサ状態のセットに基づいてDAG構造のセットを構築することと、ベイジアンネットワーク発見技術を使用することによって、グラフ構造のセットから最適なDAG構造を検索することとを備える。

30

【0009】

この実施形態の更なる変形例では、入力セットは、気象データ、ユーザ指定のブラックリスト、及びユーザ指定のホワイトリストのうちの1つ以上を含む。

【0010】

更なる変形例では、センサ状態のセットは、Kミーンズアルゴリズムを使用することによって取得される。

【0011】

この実施形態の変形例では、センサのセットの中のDAG構造は、ノードのセットを備え、グラフ構造内の各ノードは、センサクラスターのセット（すなわち、メタノード）内のセンサと、有向エッジのセットとを表し、エッジのセット内の各有向エッジは、グラフ構造内の2つのノード間の因果関係を表す。

40

【図面の簡単な説明】

【0012】

特許又は出願ファイルには、カラーで作成された少なくとも1つの図面が含まれている。本特許又は特許出願公開のカラー図面の写しは、その請求に応じて、必要な料金を支払うことにより、特許庁から提供されることとなる。

【0013】

【図1A】本発明の一実施形態に係る、例示的なデータ相関マップを示している。

50

【図 1 B】本発明の一実施形態に係る、例示的なデータ品質マップを示している。

【図 2】本発明の一実施形態に係る、例示的な根本原因分析システムアーキテクチャを示している。

【図 3 A】本発明の実施形態に係る、シミュレートされた規則的事象に対するクラスタリングトポロジーのプロットを示している。

【図 3 B】本発明の実施形態に係る、シミュレートされた不規則的事象に対するクラスタリングトポロジーのプロットを示している。

【図 4】本発明の一実施形態に係る、時間平均データに基づいて、シミュレートされたセンサデータを三元状態にマッピングする例を示している。

【図 5】本発明の一実施形態に係る、DAGモデル構造事前データモデルを示している。

10

【図 6】本発明の一実施形態に係る、図 2 のアーキテクチャについての根本原因分析システム設計を示している。

【図 7 A】本発明の一実施形態に係る、例示的な有向依存性グラフを示している。

【図 7 B】本発明の一実施形態に係る、有向依存性グラフを使用した根本原因分析の例示的な結果を示している。

【図 8】本発明の一実施形態に係る、異常事象の根本原因分析を実行するための例示的なプロセスを示すフローチャートを示している。

【図 9】本発明の一実施形態に係る、根本原因分析システムを容易にする例示的なコンピュータシステムを示している。

【図 10】本発明の一実施形態に係る、根本原因分析システムを容易にする例示的な装置を示している。

20

【0014】

図面において、同様の参照番号は、同様の図要素を指す。

【発明を実施するための形態】

【0015】

以下の説明は、当業者が実施形態を作製及び使用することを可能にするために提示され、特定の用途及びその要件に関連して提供される。開示される実施形態に対する様々な修正は、当業者には容易に明らかとなり、本明細書に定義される一般原理は、本開示の趣旨及び範囲から逸脱することなく、他の実施形態及び用途に適用され得る。したがって、本発明は、示される実施形態に限定されるものではなく、本明細書に開示される原理及び特徴と一致する最も広い範囲を与えられるものである。

30

【0016】

概要

多くの実世界業界のCPS用途では、1つ以上の機械の動作中に記録されたセンサデータは多数であり、多くの場合、信頼できない測定値を生成する不均一な種類のセンサを含む。例えば、CPS用途では、工場エネルギー管理システム(FEMS)は、例えば、毎分受信した600個のセンサからのデータなど、複数の工場のサブステーションに関連付けられた多数のセンサから周期的にデータを受信することができる。センサは、エネルギー、電流、電圧などを測定するための電気計器、温度、湿度、CO₂などを測定するための環境センサ、及び空気流、冷却設定点を測定するための制御状態センサを含むことができる。これらのセンサは、工場レイアウトにおいて様々なレベルで設置されることができる。

40

【0017】

様々なセンサからの記録されたセンサデータは、以下の不規則性を含むことができる。第1に、センサデータは、完全に異なるフォーマット及び物理的意味を有する不均質データを含むことができる。例えば、不均質データは、温度及び電力計の読み取り値を含むことができる。そのような不均質データでは、データ処理システムが、全ての変数を、根本原因分析ワークフローにおける更なる処理のために統一されたフレームワークとシームレスに変換することは困難である。第2に、センサと基地局との間の通信チャネルの不安定性に起因して、センサデータは、疎でエラーを起こしやすい可能性がある。したがって、

50

信頼性の高い正確な分析結果を生み出す一方で、全ての種類のエラーを許容することが可能なデータ処理システムが必要とされている。第3に、CPS内の多数のセンサ及び他のデータソースが、更新をデータサーバにプッシュし続けるにつれて、データサーバは、これらの更新を短時間でフラッドされる。そのような大量のデータに対処するために、データ処理システムは、データサーバ内の蓄積されたデータに対して任意の種類の分析を実行する前に、データ要約及び圧縮技術を採用することができる。

【0018】

図1Aは、本発明の一実施形態に係る、例示的なデータ相関マップを示している。データ相関マップは、FEMSからのフィーダーから収集されたデータに基づいている。図1Aでは、より明るい領域は、2つのセンサ測定値間のより高い相関を示している。一対のセンサ測定値間のそのような相関は、センサの設置位置及び種類に起因することができる。更にまた、分析されるセンサの数が増加するにつれて、より多くのセンサが相関する可能性が高い。センサデータ中のそのような高い相関の存在は、計算中に不必要なオーバーヘッドを損なう可能性がある。

【0019】

図1Bは、本発明の一実施形態に係る、例示的なデータ品質マップを示している。データ品質は、単位時間持続時間を考慮して受信されるデータサンプルの総数に対する通常のデータサンプルの比率として定義されることができる。図1Bでは、データ品質は、1時間の時間ウィンドウについて計算され、暗い領域は、通常のサンプルを示す一方で、より明るい領域は、センサデータ内の相関又はランダムな損失を示している。

【0020】

従来の自動根本原因分析方法は、センサデータ内の固有の不規則性を考慮するものではない。換言すれば、それらは、データの品質、種類及びフォーマットの仮定に基づいており、したがって、それらの自動化能力は、データセットの特性に応じて制限される。本明細書に記載される実施形態は、高次元センサデータにおける異常事象の自動根本原因分析を実行することに関連する上述の技術的問題を解決する。

【0021】

より具体的には、根本原因分析システムは、センサデータを取得し、センサデータを前処理して、サンプリングレート及び欠落したセンサ読み取り値の不規則性に対処することができる。次いで、システムは、更なる処理のために関連するセンサデータのみを選択する。具体的には、システムは、前処理されたセンサデータをコンパクトかつ情報的データ表現に変換する。更に、システムは、センサのセット内の各センサに対するセンサ状態のセットを決定する。次いで、システムは、状態のセットに基づいて、センサクラスタのセット（すなわち、メタノード）間で最適なDAG構造を構築するように努力する。DAG構造を使用することにより、システムは、ユーザ指定の標的センサに関連付けられた異常事象の根本原因を判定することができる。具体的には、システムは、直接隣接センサの状態を与えられた標的センサの異常状態の確率を判定し、次いで、標的センサにおける異常事象の根本原因は、DAG構造における異常状態をバックトラッキングすることによって判定される。

【0022】

本明細書に記載される根本原因分析システムは、以下の利点を提供する：システムが分析のために利用可能なセンサの数によって十分にスケールリングすることができるため、システムは、寸法的にスケラブルである；このシステムは、気象、時間、及びセンサデータから学習した因果関係モデルを使用して、様々な形態の異常事象の根本原因分析タスクを実行することができるために多目的に使用することができる；システムは、破損した、ノイズが多い、及び損失のあるセンサデータの分析を行うことができるためにロバストである。

【0023】

システムアーキテクチャ

図2は、本発明の一実施形態に係る、例示的な根本原因分析システムアーキテクチャを

示している。根本原因分析システム 200 は、センサデータベース 202、気象データベース 204、及び 3 つの処理ステージを含むことができる。第 1 のステージは、データ前処理モジュール 214 を含むことができ、第 2 のステージは、データ要約モジュール 220 を含むことができ、第 3 のステージは、モデル発見モジュール 228 を含むことができる。データ前処理モジュール 214 は、更に、データ取得モジュール 206、データ標準化モジュール 208、外れ値検出モジュール 210、及びデータ補間モジュール 212 を含むことができる。データ要約モジュール 220 は、更に、データ変換モジュール 216 及びセンサクラスタリングモジュール 218 を含むことができる。モデル発見モジュール 228 は、更に、自動状態分類モジュール 222、ベイジアンネットワーク構造発見モジュール 224、及び分析モジュール 226 を含むことができる。

10

【0024】

センサデータベース 202 は、工場レイアウトで 1 つ以上の機械に埋め込まれた複数のセンサから収集されたセンサデータを記憶する。センサデータは、多変量時系列として表すことができ、例えば、センサデータは、異なるタイムスタンプにおいて大きな変数群の記録を含むことができる。気象データベース 204 は、気象に関連付けられたデータ、例えば、温度、露点、湿度などを含むことができる公的に入手可能なデータベースである。データ取得モジュール 206 は、センサデータベース 202 及び気象データベース 204 を周期的に照会し、ローカル分析サーバ内のバイナリデータフォーマットに保持された対応するデータレコードを記憶及び／又は更新する。

【0025】

本発明の一実施形態では、ユーザは、全ての観察を網羅する時間間隔を示す 2 つのタイムスタンプ、すなわち、 T_s から T_e を設定することができる。例えば、時間間隔は、1 年に設定されることができる。この時間間隔は、分析期間と称されることができる。分析期間は、一対の (T_s, T_e) として表すことができ、システムは、ユーザによって明示的に指定されない限り、周期的な更新をサポートするように分析周期を自動的に設定することができる。

20

【0026】

一実施形態では、データ標準化モジュール 208 は、 T_s 及び T_e が予め定義されて変更されていないこと、及び複数のセンサに関連付けられた全ての収集された属性がインデックス

30

【0027】

【数 1】

$$i \in \{1, \dots, N\}$$

【0028】

で順序付けられることを想定することができる。ここで、 N は、分析のためにデータ取得モジュール 206 によって収集された属性の総数を表す。用語「属性」はまた、データベース内の等価用語「カラム」とも称されることができる。データ標準化モジュール 208 は、収集された属性の全てを 2 つのグループに分割することができる。第 1 のグループ

40

【0029】

【数 2】

$$S \subset \{1, \dots, N\}$$

【0030】

は、センサデータベースによって収集された属性を含む。第 2 のグループ

【0031】

【数 3】

50

$$W \subset \{1, \dots, N\}$$

【0032】

は、公共データベースからの気象関連属性を含む。更にまた、データ標準化モジュール208は、Sにおける属性を2つのクラスに分類することができる。第1のクラスは、異常事象、例えば、電力計測定値からの異常な電力消費を対象とする一次センサ

【0033】

【数4】

$$S_P \subset S$$

10

【0034】

についての属性インデックスのセットを含むことができる。第2のクラスは、S/P又はS-S_Pによって表される残りのセンサ属性を含むことができる。

【0035】

本発明の一実施形態では、データ標準化モジュール208は、カラム（又は属性）に基づいて、全ての取得されたセンサデータサンプルを時系列で変換することによって、センサデータの後続分析の効率を改善することができ、結果をカラムストアデータベースシステムに保存することができる。データ標準化モジュール208は、

20

【0036】

【数5】

$$\mathbf{a}_i = (a_{i,j})_{j=1,\dots,n_i}$$

【0037】

として各i番目の属性についての列ベクトルを構築することができ、ここで、n_iは、分析期間（T_s、T_e）中に収集されたサンプルの総数を表す。

【0038】

【数6】

$$\mathbf{t}_i = (t_{i,j})_{j=1,\dots,n_i}$$

30

【0039】

は、a_iについてのタイムスタンプの対応するアレイを示すものとし、ここで、

【0040】

【数7】

$$i \in \{1, \dots, N\} \text{ 及び } j \in \{1, \dots, n_i\}$$

40

【0041】

について、T_s ≤ t_i, j ≤ T_eである。

【0042】

センサと基地局との間のデータ通信チャネルの不安定性のために、2つの連続するタイムスタンプ間の差は、一定でないことがある。サンプリングレート内の任意の不規則性に対処するために、データ標準化モジュール208は、全てのサンプルを統一されたグローバル基準時間ドメインにマッピングすることができる。データ標準化モジュール208は、均一なグローバル基準時間を、等しい長さの間隔Δに秒で分割することができ、例えば、間隔の長さは、1時間とすることができる。

【0043】

50

【数 8】

$$\tilde{t}_{k \in \{1, \dots, K\}}$$

【0044】

は、 k 番目の時間間隔の開始時間を表し、ここで、 K は、分析期間 (T_s 、 T_e) 内の統一されたグローバル基準時間内の時間間隔の総数を表す。開始時間表記

【0045】

【数 9】

$$\tilde{t}_k$$

10

【0046】

は、センサ属性 S 及び気象属性 W の双方に使用されることができる。

【0047】

【数 10】

$$\tilde{t}_k$$

【0048】

の状態は、時間関連パラメータ表現トリプレットとして表すことができ、すなわち、 $\theta_T = (\theta_m, \theta_d, \theta_h)$ として表すことができる。トリプレットにおいて、 θ_m 、 θ_d 及び θ_h は、それぞれ、年の月、週の日、及び日の時間の略である。データ標準化モジュール 208 が、統一されたグローバル基準時間間隔に全てのタイムスタンプをマッピングした後、各属性について新しい列ベクトルを生成する。

20

【0049】

外れ値検出モジュール 210 は、各属性を 3 つの一般的な値タイプ、すなわち、連続的、離散的、及び一定の値のうちの 1 つに分類することができる。次いで、外れ値検出モジュール 210 は、列ベクトルのパターンを分析する。全てのグローバル基準時間間隔にわたって属性が一定のままである場合、これらの一定の種類の属性は、根本原因分析ワークフローにおいて可能な結果を生成しない場合があるため、更なる分析のために含まれなくてもよい。

30

【0050】

データ補間モジュール 212 は、時間間隔内でタイムスタンプを用いて全てのサンプルを集計して、数値を含む特定の属性に関連付けられた平均及び分散を計算することができる。その結果、時間間隔に少なくとも 1 つの有効なサンプルが存在する。そうでなければ、データ補間モジュール 212 は、隣接する間隔の集計結果に基づいて補間アルゴリズムを実行することができる。属性がカテゴリー値、例えば、{高、中央、低} を含む場合、データ補間モジュール 212 は、ランダムハッシュを適用して、具体的な値を整数領域にマッピングし、連続数値としてマッピング後に属性を処理することができる。例えば、データ補間モジュール 212 は、連続値タイプの線形補間、及び離散値タイプの最近傍補間を使用することによって、欠落したセンサ読み取り値を置き換えることができる。

40

【0051】

データ補間モジュール 212 の出力で利用可能なセンサデータ及び気象データに関連付けられた属性は、多様な属性タイプを含むことができる。属性タイプダイバーシティ及び多数の属性の悪影響を除去するために、データ要約モジュール 220 は、属性をコンパクトかつ情報的データ表現に変換することができる。データ要約モジュール 220 は、(データ変換モジュール 216 による) データ変換及び (センサクラスタリングモジュール 218 による) センサクラスタリングの 2 つの異なる動作を実行することができる。

【0052】

50

一実施形態では、データ変換モジュール 216 は、各時間間隔のサンプルを 2 つの明確な特徴に要約することができる；それぞれ、規則的事象及び不規則的事象の平均特徴及び分散特徴。次いで、データ変換モジュール 216 は、特徴行列を構築し、時間間隔及び属性は、行列の行及び列にそれぞれ表される。平均特徴は、周期的に発生する事象、例えば、日中又は毎週の正弦波サイクルを捕捉することができる。周期的に発生する事象は、規則的事象と称されることができる。分散特徴は、非周期的発生事象を捕捉することができ、これはまた、属性サンプル値の変動を定量化することもできる。非周期的発生事象は、不規則的事象と称されることができる。

【0053】

各時間間隔 k について、データ変換モジュール 216 は、センサデータの全ての属性に対する平均及び分散特徴を計算することができる。 $N_{i,k}$ は、時間間隔 k 中の属性 i に関連付けられたサンプルの数を表し、 $S_{i,k}$ は、時間間隔 k の開始時の属性 i のサンプルインデックス、すなわち、

【0054】

【数 11】

$$S_{i,k} = 1 + \sum_{r=1}^{k-1} n_{i,r}$$

【0055】

の開始時の属性 i のサンプルインデックスを示すものとする。そして、平均及び分散特徴は、以下のように計算される。

【0056】

【数 12】

$$\bar{a}_{i,k} = \frac{\sum_{j=S_{i,k}+1}^{S_{i,k}+n_{i,k}} a_{i,j}}{n_{i,k}} \quad (1)$$

$$\dot{a}_{i,k} = \frac{\sum_{j=S_{i,k}+1}^{S_{i,k}+n_{i,k}-1} |a_{i,j+1} - a_{i,j}|}{n_{i,k} - 1} \quad (2)$$

【0057】

ここで、

【0058】

【数 13】

$\bar{a}_{i,k}$ 及び $\dot{a}_{i,k}$

【0059】

は、時間間隔 k について属性 i の平均及び分散特徴を示す。

【0060】

センサクラスタリングモジュール 218 は、センサ測定値が非常に相関性が高いことが多く、この高い相関がランク不足の特徴行列をもたらす得ることを考慮する。したがって、類似の属性群内の有限数の代表的属性（すなわち、特徴行列の列ベクトル）を選択する必要がある。同様の属性をグループ化する前に、センサクラスタリングモジュール 218 は、グループ化／クラスタリングに対する一貫性のない単位及び属性特徴のスケールの影響を排除するために、以下の動作を実行する。最初に、センサクラスタリングモジュール 218 は、単位長さ及びゼロ中心に対する各属性の特徴ベクトル（すなわち、特徴行列の列ベクトル）を正規化することができる。次いで、センサクラスタリングモジュール 218 は、各特徴ベクトルを単位 K 次元球体の表面上に等価にマッピングすることができる。

【 0 0 6 1 】

本発明の一実施形態では、センサクラスタリングモジュール 2 1 8 は、親和性クラスタリングアルゴリズムを使用することができる。親和性クラスタリングアルゴリズムを使用して、各群が、例示的なものと呼ぶことができる対応する単一の代表的属性を有し、グループ内の残りの属性を従動するものと呼ぶことができるような方法でグループ化することができる。各例示的なものの場合、センサクラスタリングモジュール 2 1 8 は、2 つの特徴行列、すなわち、平均特徴行列及び変動特徴行列を構築することができる。データセンサクラスタリングモジュール 2 1 8 で実行されるデータ処理動作は、多数の相関属性を排除することによって、モデル発見モジュール 2 2 8 内の探索空間を低減することができる。更にまた、親和性クラスタリングアルゴリズムは、データポイント間の距離測定基準を通過及び使用するメッセージを介してクラスタを構築することによって、大きなデータセットを処理するための高速かつ信頼性の高い方法を提供する。

10

【 0 0 6 2 】

以下のセクションでは、センサクラスタリングモジュール 2 1 8 の動作について、更に詳細に説明する。本発明の一実施形態では、センサクラスタリングモジュール 2 1 8 は、それぞれ、式 (1) 及び (2) に従って属性 i の平均及び分散特徴ベクトルを計算することができる。平均特徴ベクトルは、

【 0 0 6 3 】

【数 1 4】

$$\bar{\mathbf{a}}_i = (\bar{a}_{i,1}, \dots, \bar{a}_{i,K})$$

20

【 0 0 6 4 】

として表すことができ、分散特徴ベクトルは、

【 0 0 6 5 】

【数 1 5】

$$\dot{\mathbf{a}}_i = (\dot{a}_{i,1}, \dots, \dot{a}_{i,K})$$

【 0 0 6 6 】

として表すことができる。平均特徴ベクトル

30

【 0 0 6 7 】

【数 1 6】

$$\bar{\mathbf{a}}_i$$

【 0 0 6 8 】

は、

【 0 0 6 9 】

【数 1 7】

$$\bar{\mathbf{x}}_i = (\bar{\mathbf{a}}_i - \bar{\bar{\mathbf{a}}}_i) / \|\bar{\mathbf{a}}_i - \bar{\bar{\mathbf{a}}}_i\|_{l_2}$$

40

【 0 0 7 0 】

であるように、

【 0 0 7 1 】

【数 1 8】

$$\bar{\mathbf{x}}_i$$

【 0 0 7 2 】

によって示されるゼロ中心及び単位長さベクトルに変換されることができ、ここで、

50

【 0 0 7 3 】

【数 1 9】

$$\bar{\bar{\mathbf{a}}}_i = \sum_k \mathbf{a}_{i,k} / K$$

【 0 0 7 4 】

である。同様に、分散特徴ベクトル

【 0 0 7 5 】

【数 2 0】

$$\dot{\mathbf{a}}_i$$

10

【 0 0 7 6 】

も変換されることができる。属性 i と j との間の平均特徴における相違性は、

【 0 0 7 7 】

【数 2 1】

$$d_{ij}^{avg} = |\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j|_{l_2}$$

【 0 0 7 8 】

としてユークリッド距離によって測定することができる。同様に、分散特徴ベクトル

【 0 0 7 9 】

【数 2 2】

$$\dot{\mathbf{a}}_i$$

20

【 0 0 8 0 】

における相違度を測定することができ、

【 0 0 8 1 】

【数 2 3】

$$d_{ij}^{var}$$

30

【 0 0 8 2 】

として示される。

【 0 0 8 3 】

平均特徴ベクトルにおける相違度はまた、

【 0 0 8 4 】

【数 2 4】

$$d_{ij}^{avg} = \sqrt{2(1 - \rho_{ij}^{avg})}$$

40

【 0 0 8 5 】

として定義することができ、ここで、

【 0 0 8 6 】

【数 2 5】

$$\rho_{ij}^{avg}$$

50

【 0 0 8 7 】

は、属性 i と j との間の平均特徴の相関係数である。更に、平均特徴ベクトルの相違度尺度は、実数

【 0 0 8 8 】

【数 2 6】

$$0 \leq d_{ij}^{avg} \leq 2$$

【 0 0 8 9 】

のセットによって境界付けられ、平均特徴ベクトルの類似度は、

10

【 0 0 9 0 】

【数 2 7】

$$s_{ij}^{avg} = 2 - d_{ij}^{avg}$$

【 0 0 9 1 】

として定義することができる。分散特徴ベクトルの類似度は、

【 0 0 9 2 】

【数 2 8】

$$s_{ij}^{var} = 2 - d_{ij}^{var}$$

20

【 0 0 9 3 】

として定義することができる。

【 0 0 9 4 】

そして、センサクラスタリングモジュール 2 1 8 は、以下の条件が満たされた場合に、平均及び分散特徴ベクトルのゼロ分散属性を決定することができる。

【 0 0 9 5 】

【数 2 9】

$$Var(\bar{x}_i) \leq \delta_{avg} \quad (3)$$

30

$$Var(\dot{x}_i) \leq \delta_{var} \quad (4)$$

【 0 0 9 6 】

ここで、

【 0 0 9 7 】

【数 3 0】

$$Var(\bar{x}_i)$$

40

【 0 0 9 8 】

は、正規化された平均特徴ベクトル及び

【 0 0 9 9 】

【数 3 1】

$$Var(\dot{x}_i)$$

【 0 1 0 0 】

は、正規化された分散特徴ベクトルの分散を示し、

【 0 1 0 1 】

50

【数 3 2】

δ_{avg} 及び δ_{var}

【0 1 0 2】

は、

【0 1 0 3】

【数 3 3】

$\delta_{avg} = \delta_{var} = 10^{-2}$

10

【0 1 0 4】

でのそれぞれの閾値である。

【0 1 0 5】

【数 3 4】

C_{avg}^{zero} 及び C_{var}^{zero}

【0 1 0 6】

を、平均及び分散特徴ベクトルのゼロ分散属性のインデックスのグループ（又はクラス）をそれぞれ示すものとする。一実施形態では、センサクラスタリングモジュール 218 は、親和性クラスタリングアルゴリズムを使用して、平均特徴及び分散特徴のクラスを計算することができる。親和性クラスタリングアルゴリズムは、それぞれ、全ての非ゼロ分散属性

20

【0 1 0 7】

【数 3 5】

$i, j \notin C_{avg}^{zero}$ 及び $i, j \notin C_{var}^{zero}$

【0 1 0 8】

について、類似度が

【0 1 0 9】

【数 3 6】

s_{ij}^{avg} 及び s_{ij}^{var}

30

【0 1 1 0】

を使用する。そして、例示的な属性の平均及び分散特徴は、それぞれ、特徴行列

【0 1 1 1】

【数 3 7】

$\mathbf{X}^{avg} = [\bar{\mathbf{x}}_i]_{i \in C_{avg}^{exmp}}$ 及び $\mathbf{X}^{var} = [\dot{\mathbf{x}}_i]_{i \in C_{var}^{exmp}}$

40

【0 1 1 2】

（式中、

【0 1 1 3】

【数 3 8】

C_{avg}^{exmp} 及び C_{var}^{exmp}

50

【0114】

は、平均及び分散特徴の例示の属性指数である。

【0115】

図3Aは、本発明の一実施形態に係る、シミュレートされた規則的事象についての例示的なもの及び従動するもののセンサのクラスタリングトポロジーのプロットを示している。図3Bは、本発明の一実施形態に係る、シミュレートされた不規則的事象についての例示的なもの及び従動するもののセンサのクラスタリングトポロジーのプロットを示している。図3A及び図3Bのトポロジーは、シミュレートされた規則的及び不規則的事象についての親和性クラスタリングアルゴリズムを使用することによって得られている。

【0116】

図2に戻ると、モデル発見モジュール228は、以下のデータ処理モジュール：自動状態分類モジュール222、ベイジアンネットワーク構造発見モジュール224及び分析モジュール226を使用することによって、センサクラスタリングモジュール218によって出力されるデータを更に処理する。モデル発見モジュール228の目的は、観察されたセンサデータ及び気象データを所与として最適なベイジアンネットワーク（BN）構造を検索することによって、属性間の因果関係の単純なグラフ表示をDAGとして見つけることである。

【0117】

自動状態分類モジュール222は、ニーポイント検出によるKミーンズアルゴリズムを使用することができ、最適数のクラスタKは、KとKを与えられたモデルの対数尤度との間のトレードオフのニーポイントによって判定されることができる。それにより、三元状態、例えば、高、低、及び非ピーク、及び二元状態、例えば、それぞれ規則的及び不規則的変動によって表される、規則的事象及び不規則的事象を伴うBN構造の状態構成を簡略化する。図4は、本発明の一実施形態に係る、シミュレートされたセンサデータの時間平均に基づいて、シミュレートされたセンサデータを三元状態にマッピングする例を示している。図4のプロット（a）は、目標センサ測定値、すなわち、電力計データの電力消費プロファイルの例を示している。この電力計データを最初に15分の平均に変換することができ、そのような平均電力計データをプロット（b）に示している。次いで、自動状態分類モジュール222は、平均電力測定器データを三元状態、例えば、高、低、及び非ピークに変換することができる（プロット（c）に示す）。

【0118】

表1（以下に示す）は、自動状態分類モジュール222によって実行されるセンサ測定の事象分類を示している。表1はまた、平均及び変動特徴の特性間の比較を提供する。

【0119】

【表1】

表1：センサ測定値の事象分類

事象種類	規則的事象（平均値）	不規則的事象（微分値）
特性	周期的	非周期的
時間ウィンドウ	ユーザによって設定（デフォルト15分）	
信号特徴	サンプル平均	総変形
状態分類	三元状態 （低、高、又は非ピーク）	二元状態 （規則的又は不規則的変動）
方法	ニーポイント検出を有するKミーン	

【0120】

図2に戻ると、ベイジアンネットワーク構造発見モジュール224は、次に、ノード間の有向エッジによってDAGとして表される属性間の因果関係を伴う特徴行列に基づいて、最も確率の高いグラフィック構造のセットを見つけることができる。グラフィック構造のそれぞれは、センサに関連付けられた属性に対応する各ノードを有するノードのセットと、ノード間の有向エッジのセットと、を含むことができる。

【 0 1 2 1 】

以下のセクションでは、ベイジアンネットワーク構造発見モジュール 2 2 4 及び分析モジュール 2 2 6 の動作について、更に詳細に説明する。従来のシステムは、最適な B N 構造を特定するための指数関数的探索空間を探索する必要がある、これは、B N 構造及び状態に関する任意の制約なしに最適な B N 構造を同定する必要がある。従来のシステムが、最適な B N 構造を有限時間で見つけることができる場合であっても、得られる構造は、初期出発点に対して高度に敏感であり、したがって信頼できない。更にまた、B N 構造中の状態が正確に定義されない場合、B N 構造を学習するために使用されるアルゴリズムは、無限に遅い速度で最終解に収束し得る。

【 0 1 2 2 】

上述の問題を克服するために、本発明の一実施形態は、探索を誘導するために、B N 構造の従来のデータモデルを設定し、図 5 は、そのような事前データモデルを示している。図 5 は、本発明の一実施形態に係る、D A G モデル構造事前データモデルを示している。事前データモデルは、B N 構造上の気象属性 5 0 2 の時間パラメータ 5 0 6 及びセンサ属性（標的センサ測定 5 0 8 及び非標的センサ測定 5 0 4 を含む）の間の因果関係の基本的な制約を課す。因果関係の 1 つの方向のみが、例えば、時間パラメータ 5 0 6 から気象属性 5 0 2 に許可されるが、気象属性 5 0 2 から時間パラメータ 5 0 6 までは許可されない。そのような因果的制約は、任意の種類のセンサ及び気象データに対して普遍的に保持された基本的な B N 構造を定義することができる。システムは、図 2 の自動状態分類モジュール 2 2 2 によって学習された個々の属性の隠れたパラメータを使用することによって因果関係モデルを表すことができる。一実施形態では、図 2 の自動状態分類モジュール 2 2 2 は、連続状態の最適な量子化を見つけることと同等であり得る、属性の特徴の連続状態のために、最適な離散状態、すなわち隠れパラメータを見出すことができる。次いで、ユーザは、それらの既知の依存知識に基づいて、ブラック／ホワイトリスト 5 1 0 を提供して、B N 構造を更に制約することができる。

【 0 1 2 3 】

一実施形態では、表記 X は、 X^{avg} 及び X^{var} の双方を表すために使用される。表記 X_i は、 X の i 番目の列ベクトルを示し、属性 i の観察された特徴サンプルを表す。標的センサ測定属性のための特徴行列は、

【 0 1 2 4 】

【数 3 9】

$$X^{SP}、及び X^{SP} = (X_i)_{i \in SP}$$

【 0 1 2 5 】

で表すことができる。同様に、非標的センサ測定属性 5 0 4 及び気象属性 5 0 2 についての

【 0 1 2 6 】

【数 4 0】

$$X^{S/P}$$

【 0 1 2 7 】

及び X^W をそれぞれ定義することができる。一実施形態では、観察された特徴サンプル $X_{i,k}$ は、下にある又は隠れた離散した状態 θ_i 、すなわち、

【 0 1 2 8 】

【数 4 1】

$$P(x_{i,k} | X, (\theta_j)_{j=1, \dots, N}) = P(x_{i,k} | \theta_i)$$

【 0 1 2 9 】

に依存し得る。あるいは、 θ_i は、量子化後の連続ランダム変数 $X_{i,k}$ の離散状態表現として解釈することができる。表記

【 0 1 3 0 】

【数 4 2】

θ^{SP}

【 0 1 3 1 】

は、

【 0 1 3 2 】

【数 4 3】

X^{SP}

10

【 0 1 3 3 】

の隠れた状態を示し、

【 0 1 3 4 】

【数 4 4】

$\theta^{S/P}$

【 0 1 3 5 】

は、

【 0 1 3 6 】

【数 4 5】

$X^{S/P}$

20

【 0 1 3 7 】

の隠れた状態を示し、同様に、気象及び時間属性の隠れた状態は、それぞれ θ^W 及び θ^T によって示され得る。属性間の因果関係は、図 5 に示す DAG を使用することによって、隠れパラメータ

【 0 1 3 8 】

【数 4 6】

$\theta = (\theta^{SP}, \theta^{S/P}, \theta^W, \theta^T)$

30

【 0 1 3 9 】

によってコンパクトに表すことができる。図 5 では、ノード間の矢印は因果関係を表し、各プレートは、同じ属性を有するノード間の一例のみを示す。プレート内のノード間の依存性構造については、更なる想定はなされていない。

【 0 1 4 0 】

図 2 に戻ると、ベイジアンネットワーク構造発見モジュール 224 は、図 5 のグラフィックモデル及び特徴行列 X を考慮して、

【 0 1 4 1 】

【数 4 7】

$P(\theta^{SP} | \theta^{S/P}, \theta^W, \theta^T)$

40

【 0 1 4 2 】

を見つけるように、

【 0 1 4 3 】

【数 4 8】

50

θ^{Sp}

【0144】

コンパクトな確率的説明を見つける。この目的は、図5のX及びモデルを考慮して、

【0145】

【数49】

 θ^{Sp}

【0146】

について最良の図形構造を体系的に検索することを容易にするBNフレームワークを採用することによって達成することができる。

【0147】

 $G = (N, E)$ は、ノード

【0148】

【数50】

 $N = \{\theta_i \in \Theta\}$

【0149】

を有するDAG、及び変数間の直接的な依存性を表すエッジEを示す。更に、更なる変数は、以下のように定義することができる。

【表2】

変数	定義
Π_{θ_i}	G中の θ_i の親の集合
$D_{\Pi_{\theta_i}}$	Π_{θ_i} における変数のジョイントドメイン
θ_k	θ のK番目の値
w_{ij}	Π_{θ_i} のj番目の構成
z_{ijk}	$z_{ijk} = P(\theta_i = \theta_{ik} \Pi_{\theta_i} = w_{ij})$
Z	$Z = \{z_{ijk}\}_{i \in 1 \dots n, j \in D_{\Pi_{\theta_i}}, k \in D_i}$

【0150】

n次元のBNは、トリプレット $B = (\theta, G, Z)$ と、以下によって与えられる θ に対する固有のジョイント確率分布として定義することができる：

【0151】

【数51】

$$P_B(\theta_1, \dots, \theta_n) = \prod_{i=1}^n P_B(\theta_i | \Pi_{\theta_i}) \quad (5)$$

【0152】

n個の変数を有する全てのベイジアンネットワークのセットは、

【0153】

【数52】

 $\mathcal{B}_n$

【0154】

によって表すことができる。一実施形態では、ベイジアンネットワーク構造発見モジュ

10

20

30

40

50

ール 2 2 4 は、値

【 0 1 5 5 】

【数 5 3】

$\phi(B, \mathcal{D}_\theta)$

【 0 1 5 6 】

を最大化するベイジアンネットワーク

【 0 1 5 7 】

【数 5 4】

$B \in \mathcal{B}_n$

10

【 0 1 5 8 】

を探索することにより、ベイジアンネットワークを学習することができ、ここで、 ϕ はスコアリング機能を示し、

【 0 1 5 9 】

【数 5 5】

$\mathcal{D}_\theta$

20

【 0 1 6 0 】

は、 X から見出される隠れ状態 θ のサンプルを示す。一実施形態では、ベイジアンネットワーク構造発見モジュール 2 2 4 は、ベイジアン情報基準 (B I C) スコアによって

【 0 1 6 1 】

【数 5 6】

$\mathcal{B}_n$

【 0 1 6 2 】

を検索するための G r e e d y のヒルクライミング法を使用することができる。

【 0 1 6 3 】

ベイジアンネットワーク構造発見モジュール 2 2 4 は、ベイジアンネットワークフレームワークを使用して、不規則事象及び規則的事象に対する依存性モデルを説明することができる。図 5 に示す基本依存性モデルを考慮すると、ベイジアンネットワーク構造発見モジュール 2 2 4 は、不規則的及び規則的事象のために、完全な依存性構造 (B N 構造) を自動的に学習することができる。次いで、完全な依存性構造を所与として、ベイジアンネットワークパラメータ推定アルゴリズムを使用して、規則的及び不規則的事象に対する依存確率 (すなわち、B N 分布) を計算することができる。本発明の一実施形態では、ベイジアンネットワーク構造発見モジュール 2 2 4 は、センサ測定値、気象データ (例えば、温度、露点、湿度、イベント、及び条件)、及びタイムスタンプ (月、週、日、時間) から、B N を学習することができる。更に、システムはまた、ユーザが入力 (例えば、ブラックリスト及びホワイトリスト) を図 5 の学習モデルに手動で提供することを可能にすることができる。

30

40

【 0 1 6 4 】

分析モジュール 2 2 6 は、次に、ノード i がリーフノード (すなわち、子なしのノード) であるまで、以下の動的方程式 (6) を連続的に計算することによって、根本原因分析を実行する。

【 0 1 6 5 】

【数 5 7】

50

$$\theta_{j_i}^k \leftarrow \arg \max_{\theta_{j_i}} P_B(\theta_i = \theta_{j_i}^k | \theta_{j_i}), \quad \theta_{j_i}^0 \leftarrow \theta_t^* \quad (6)$$

【0166】

ここで、 j_i はノード i の親ノードを示し、 θ_{j_i} は、ノード i の親ノードの状態を表し

、
【0167】

【数58】

θ_t^*

10

【0168】

は標的センサの異常状態を示す。

【0169】

図6は、本発明の一実施形態に係る、図2のアーキテクチャについての根本原因分析システム設計を示している。図6は、2つの独立して実行されるシステム：トレーニングのためのシステム612、及び推論のためのシステム614を示している。一実施形態では、トレーニングシステム612は、データベースサーバからのデータストリーム602のバッチからベイジアンネットワーク構造606を学習し、工場レイアウト618の事前知識を学習することができる。推論システム614は、トレーニングデータセット608で利用可能でありかつユーザ定義の標的センサ616が与えられたベイジアンネットワーク構造の最新の更新を使用して、根本原因分析タスク610を実行することができる。メッセージ待ち行列604は、データベース、トレーニングシステム612、及び推論システム614の間のデータストリームを適切に制御することができる。

20

【0170】

図7Aは、本発明の一実施形態に係る、例示的な有向依存性グラフを示している。ノード702-732は、異なるセンサを表し、ノードを接続する有向エッジは、ノード間の接続を表す。図7Bは、本発明の一実施形態に係る、有向依存性グラフを使用した根本原因分析の例示的な結果を示している。標的センサ710（黒影付き円）が与えられると、システムは、図7Bに示される有向依存性グラフを使用して、根本原因分析を実行する。一実施形態では、根本原因分析は、式(6)を使用して計算される。例えば、所与の標的センサ710の異常状態が「アイドル」状態に設定されている場合、式(6)を使用して、以下の機械に存在する条件により、センサ710内の「アイドル」状態が最も起こりやすいと結論付けることができる：センサ720に関連付けられた機械もアイドル状態である場合、センサ716に関連付けられた機械の入力において利用可能な入力がない場合、及びセンサ712に関連付けられた機械の接触時間が通常よりも長い場合。また、異常の最も確率が高い原因は、その状態の最も確率の高い原因がセンサ716及び730のうちのセンサ730'のオフ状態である、センサ716の「入力なし」状態によるものである。したがって、所与の標的センサ710内の異常状態の根本原因は、センサ730の特定のオフ状態にトレースバックされることができる。

30

40

【0171】

図8は、本発明の一実施形態に係る、異常事象の根本原因分析を実行するための例示的なプロセスを示すフローチャートを示している。動作中、システムは、センサデータベースに記憶されたセンサデータを取得することができる（動作802）。センサデータは、工場レイアウトで1つ以上の機械に埋め込まれた複数のセンサに関連付けられる。

【0172】

産業用IoT用途では、センサの数の増加に起因して、収集されたセンサデータの量は大きくなり得る。更にまた、センサデータは、通常、欠落した、破損した、ノイズが多い、及び非常に相関がある値を含む、多数の悪状態データを含む。このような不良な品質のセンサデータが根本原因分析を実行するために使用される場合、結果は不正確であり、し

50

たがって信頼することができない。更にまた、根本原因分析の計算複雑性は、多数のセンサデータに起因して大幅に増加し得る。センサデータのコンパクトかつ情報的表現を提供するために、システムは、センサデータをセンサ状態のセットに変換する（動作 804）。

【0173】

動作 804 に続いて、システムは、センサ状態のセットを使用して、最適な DAG 構造を連続的に学習し、最も確率の高い BN 構造を選択することによって、最適な BN 構造を構築することができる（動作 806）。次いで、システムは、最適な DAG 構造について、グラフ内の直接隣接するもののうちのいくつかの状態を所与として、ユーザ指定の標的センサの異常状態の確率を判定することができる（動作 808）。次いで、システムは、DAG 上の異常状態をバックトラッキングすることによって、標的センサの異常状態の最も確率可能な原因を判定することができる（動作 810）。

10

【0174】

例示的なコンピュータシステム及び装置

図 9 は、本発明の一実施形態に係る、根本原因分析システムを容易にする例示的なコンピュータシステムを示している。コンピュータシステム 900 は、プロセッサ 902、メモリ 904、及び記憶デバイス 906 を含む。コンピュータシステム 900 は、ディスプレイ装置 910、キーボード 912、及びポインティングデバイス 914 に結合されることができ、また、1 つ以上のネットワークインターフェースを介してネットワーク 908 に結合されることができる。記憶装置 906 は、オペレーティングシステム 918 及び根本原因分析システム 920 を記憶することができる。

20

【0175】

根本原因分析システム 920 は、コンピュータシステム 900 によって実行されると、コンピュータシステム 900 に、本開示で説明される方法及び／又はプロセスを実行させることができる命令を含むことができる。根本原因分析システム 920 はまた、1 つ以上のセンサに関連付けられたセンサデータを受信するための命令（センサデータ受信モジュール 922）、センサデータ前処理を実行するための命令（センサデータ前処理モジュール 924）、及びデータ前処理ステージの後にセンサデータを更に処理するための命令（データ要約モジュール 926）と、を含むことができる。更にまた、根本原因分析システム 920 は、最も確率の高いグラフィカル構造を発見し、ユーザ指定の標的センサにおける異常の根本原因を判定するための命令（モデル発見モジュール 928）を含むことができる。

30

【0176】

図 10 は、本発明の一実施形態に係る、監視されていない異常検出システムを容易にする例示的な装置を示している。装置 1000 は、有線、無線、量子光、又は電気通信チャネルを介して互いに通信し得る複数のユニット又は装置を備えることができる。装置 1000 は、1 つ以上の集積回路を使用して実現されることができ、図 10 に示されているものよりも少ない又は多いユニット又は装置を含むことができる。更に、装置 1000 は、コンピュータシステムに統合され得るか、又は他のコンピュータシステム及び／若しくはデバイスと通信することができる別個のデバイスとして実現され得る。具体的には、装置 1000 は、図 9 のコンピュータシステム 900 のモジュール 920-928 と同様の機能又は動作を実行するユニット 1002-1010 を備えることができ、センサデータ受信ユニット 1002 と、センサデータ前処理ユニット 1004 と、データ要約ユニット 1006 と、モデル発見ユニット 1008 とを含む。装置 1000 は、通信ユニット 1010 を更に含むことができる。

40

【0177】

一般に、本発明の実施形態は、DAG を使用して高次元センサデータの異常事象の根本原因を追跡する方法及びシステムを提供する。センサデータを使用して工場機械動作における異常の根本原因を発見することが、例として使用されている。実際には、この解決策は、工場機械に関連付けられた異常の検出に限定されるものではない。また、他の種類の装置又は機械装置における異常を検出するために使用することもできる。

50

【0178】

「発明を実施するための形態」の節に記載される方法及びプロセスは、上に論じられるようなコンピュータ可読記憶媒体内に記憶され得るコード及び／又はデータとして具体化され得る。コンピュータシステムが、コンピュータ可読記憶媒体上に記憶されたコード及び／又はデータを読み取って実行すると、コンピュータシステムは、データ構造及びコードとして具体化され、コンピュータ可読記憶媒体内に記憶された方法及び処理を実行する。

【0179】

更に、上述の方法及び処理は、ハードウェアモジュール又は装置に含まれてもよい。ハードウェアモジュール又は装置としては、特定用途向け集積回路（*application-specific integrated circuit*、*ASIC*）チップ、フィールドプログラム可能ゲートアレイ（*field-programmable gate array*、*FPGA*）、特定の時刻に特定のソフトウェアモジュール又はコードを実行する専用又は共有プロセッサ、及び、既知の又は後に開発される他のプログラム可能論理デバイスを含むことができるが、これらに限定されない。ハードウェアモジュール又は装置が起動されると、それらの内部に含まれる方法及び処理が実行される。

【0180】

本明細書に記載される前述の実施形態は、例示及び説明のみを目的として提示されている。それらは、網羅的であること、又は本発明を開示される形態に限定することを意図するものではない。したがって、多くの修正及び変形が、当業者には明らかであろう。更に、上記の開示は、本発明を限定することを意図するものではない。本発明の範囲は、添付の特許請求の範囲によって定義される。

10

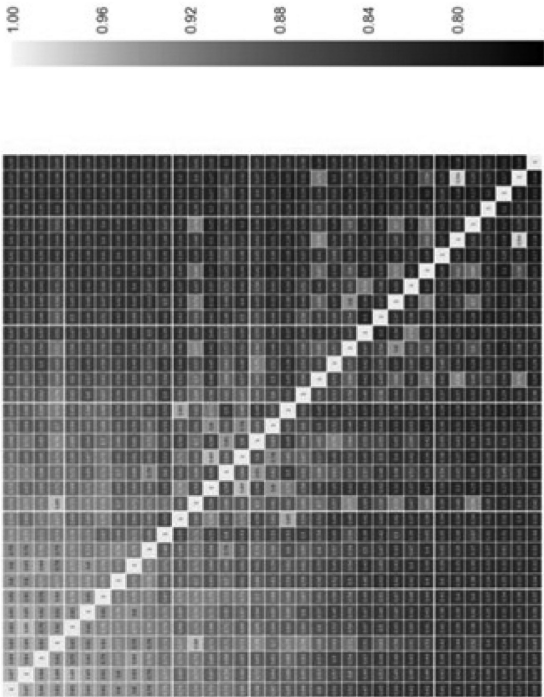
20

30

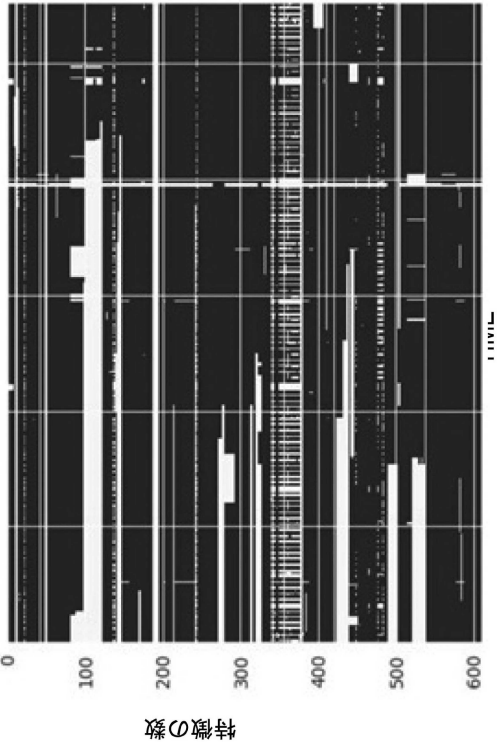
40

50

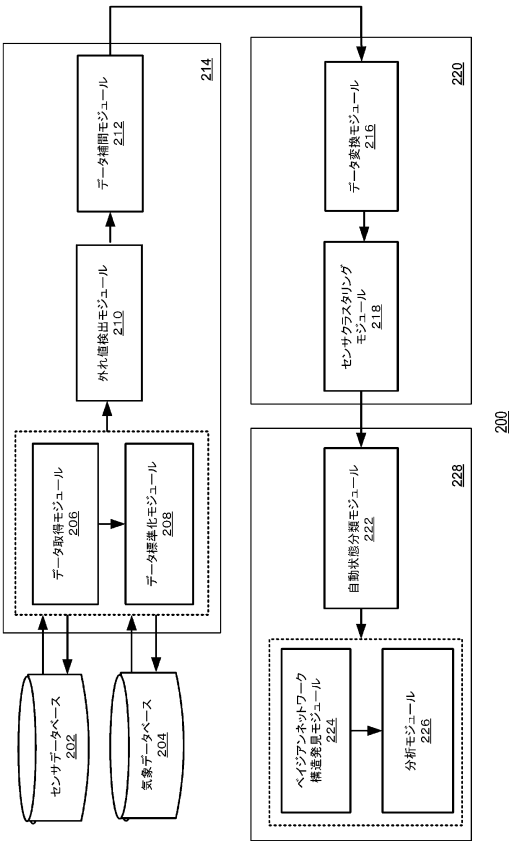
【図面】
【図 1 A】



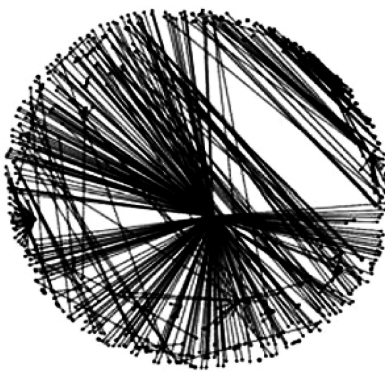
【図 1 B】



【図 2】



【図 3 A】



10

20

30

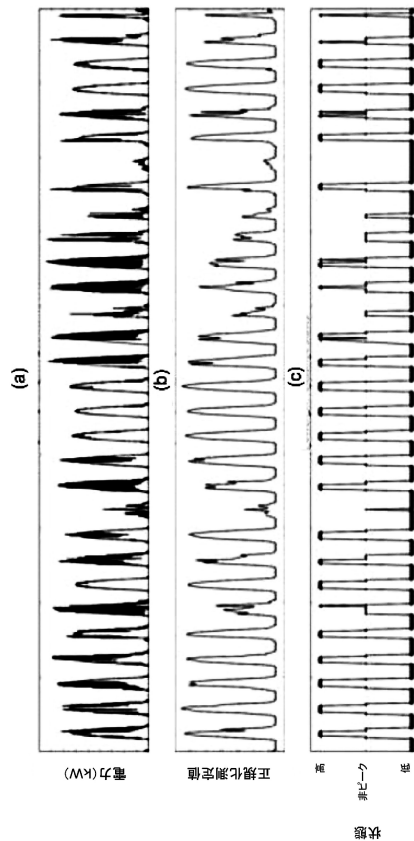
40

50

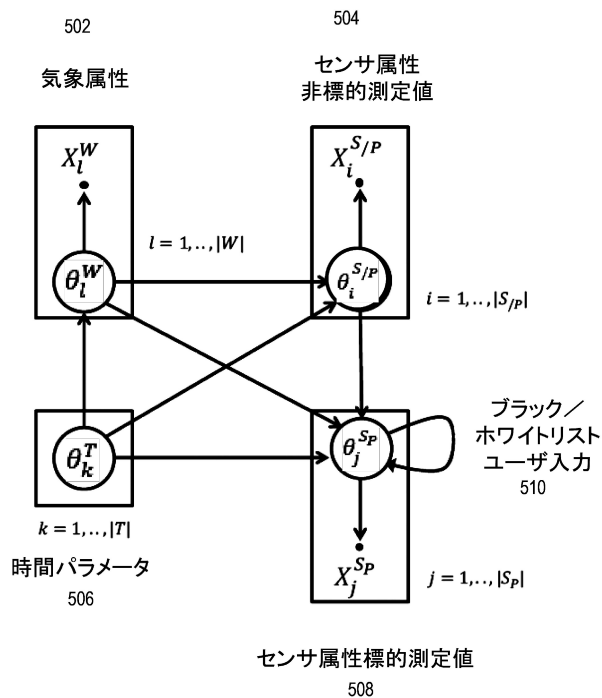
【図 3 B】



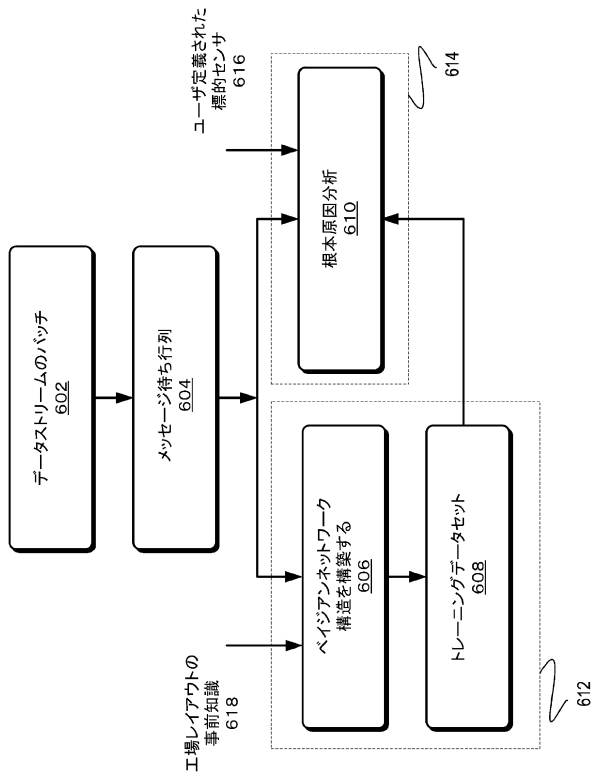
【図 4】



【図 5】



【図 6】



10

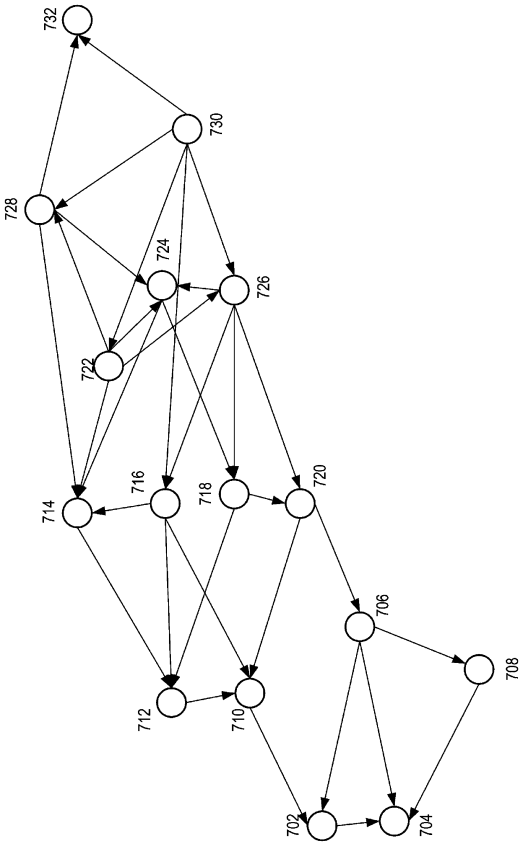
20

30

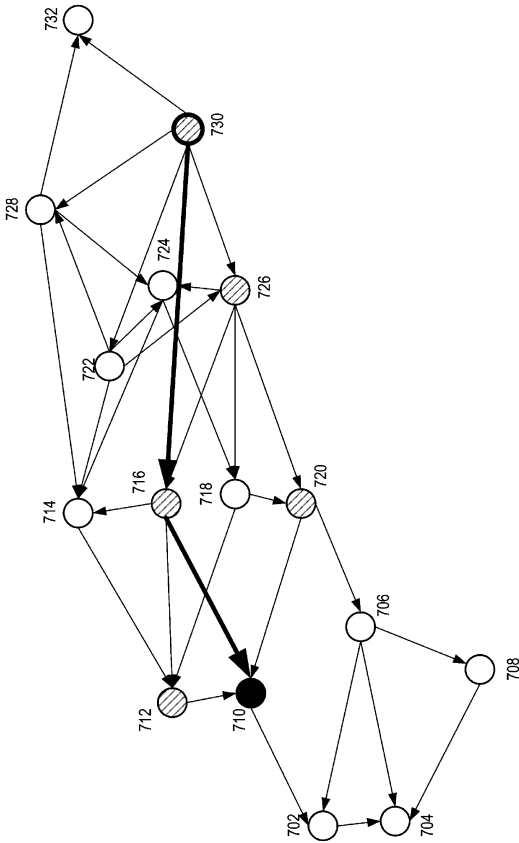
40

50

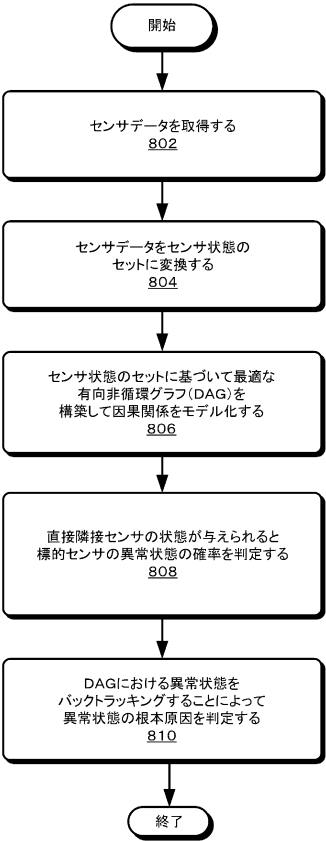
【図 7 A】



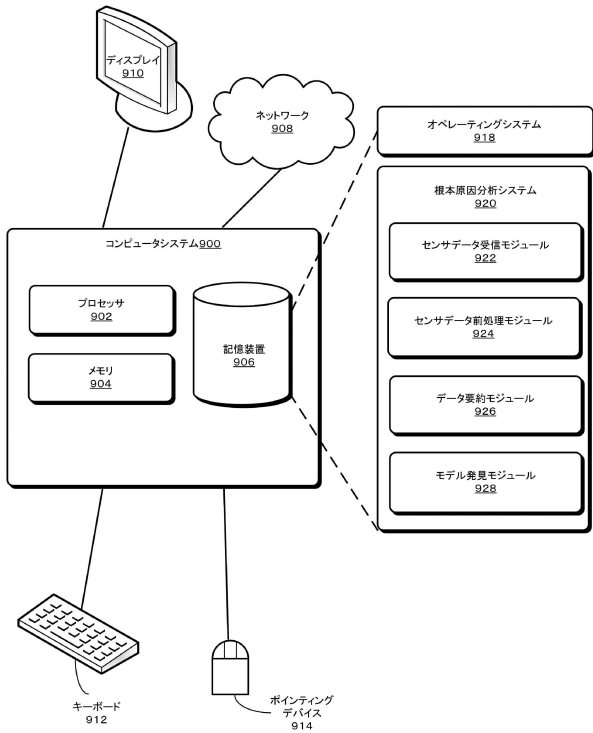
【図 7 B】



【図 8】



【図 9】



10

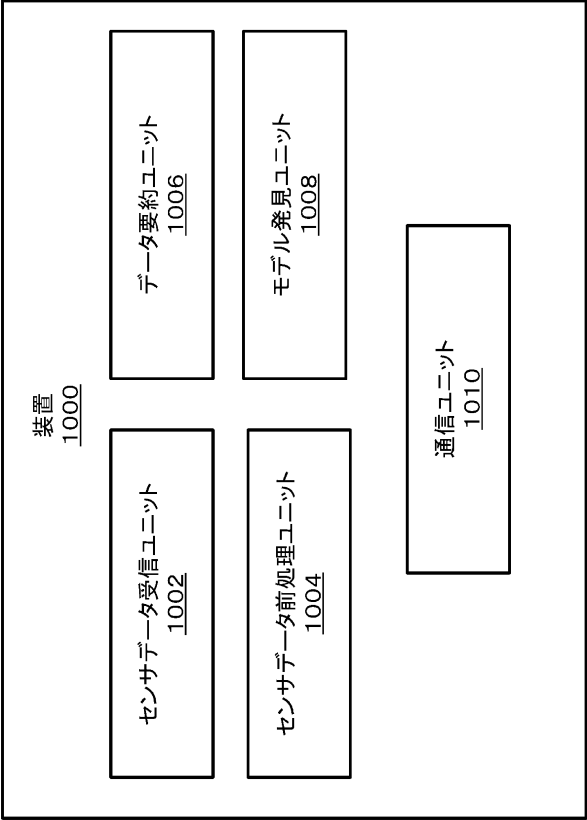
20

30

40

50

【図 10】



10

20

30

40

50

フロントページの続き

(51)国際特許分類

F I
G 0 6 F 11/07 1 9 0

(74)代理人 弁理士 上杉 浩

100120525

(74)代理人 弁理士 近藤 直樹

100139712

(74)代理人 弁理士 那須 威夫

100158551

(72)発明者 弁理士 山崎 貴明

ドクウー・ジュン

アメリカ合衆国 カリフォルニア州 9 4 0 4 0 マウンテン・ビュー サウス・レンジストルフ・
アベニュー 5 7 5 ナンバー 1 3 2

審査官 山本 俊介

(56)参考文献 米国特許出願公開第 2 0 1 5 / 2 8 0 9 6 9 (U S , A 1)

米国特許出願公開第 2 0 1 8 / 2 1 1 1 7 1 (U S , A 1)

Sarah Nadi, DRACA: Decision-support for Root Cause Analysis and Change Impact Analysis
, UWSpace [online], 2009年, https://uwspace.uwaterloo.ca/bitstream/handle/10012/4889/NADI_SARAH.pdf, [2023年8月15日検索]

(58)調査した分野 (Int.Cl., D B 名)

G 0 6 N 7 / 0 1

G 0 5 B 1 9 / 4 1 8

G 0 5 B 2 3 / 0 2

G 0 6 F 1 1 / 0 7