

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7534698号
(P7534698)

(45)発行日 令和6年8月15日(2024.8.15)

(24)登録日 令和6年8月6日(2024.8.6)

(51)国際特許分類

F I

G 0 6 N 7/01 (2023.01)

G 0 6 N 7/01

請求項の数 7 (全9頁)

(21)出願番号	特願2023-523709(P2023-523709)	(73)特許権者	000004226
(86)(22)出願日	令和3年5月24日(2021.5.24)		日本電信電話株式会社
(86)国際出願番号	PCT/JP2021/019524		東京都千代田区大手町一丁目5番1号
(87)国際公開番号	WO2022/249224	(74)代理人	100083806
(87)国際公開日	令和4年12月1日(2022.12.1)		弁理士 三好 秀和
審査請求日	令和5年10月23日(2023.10.23)	(74)代理人	100129230
			弁理士 工藤 理恵
		(72)発明者	佐藤 亮介
			東京都千代田区大手町一丁目5番1号
			日本電信電話株式会社内
		(72)発明者	竹下 恵
			東京都千代田区大手町一丁目5番1号
			日本電信電話株式会社内
		(72)発明者	高田 篤
			東京都千代田区大手町一丁目5番1号
			最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57)【特許請求の範囲】

【請求項1】

ベイジアンネットワークにおけるデータの誤りを検出する情報処理装置であって、
投入したデータに基づくベイジアンネットワークの各ノードの判断の傾向についてケン
ドールの一致度係数を計算する計算部と、
前記kendールの一致度係数が閾値より低い場合に、前記データが誤りを含む旨の判定
結果を出力する出力部を備える
情報処理装置。

【請求項2】

請求項1に記載の情報処理装置であって、
前記計算部は、前記投入したデータに基づきベイジアンネットワークの各ノードについ
て子ノードの事後確率値のランキングを求め、求めたランキングについてkendールの一
致度係数を計算する
情報処理装置。

【請求項3】

請求項1または2に記載の情報処理装置であって、
前記出力部は、前記kendールの一致度係数が閾値より低い場合に、判断の傾向が他と
異なるノードに投入された前記データが誤りを含む旨の判定結果を出力する
情報処理装置。

【請求項4】

ベイジアンネットワークにおけるデータの誤りを検出する情報処理方法であって、
コンピュータが、
投入したデータに基づくベイジアンネットワークの各ノードの判断の傾向についてケンドールの一致度係数を計算し、
前記ケンドールの一致度係数が閾値より低い場合に、前記データが誤りを含む旨の判定結果を出力する
情報処理方法。

【請求項 5】

請求項 4 に記載の情報処理方法であって、
前記投入したデータに基づきベイジアンネットワークの各ノードについて子ノードの事後確率値のランキングを求め、求めたランキングについてケンドールの一致度係数を計算する
情報処理方法。

10

【請求項 6】

請求項 4 または 5 に記載の情報処理方法であって、
前記ケンドールの一致度係数が閾値より低い場合に、判断の傾向が他と異なるノードに投入された前記データが誤りを含む旨の判定結果を出力する
情報処理方法。

【請求項 7】

請求項 1 ないし 3 のいずれかに記載の情報処理装置の各部としてコンピュータを動作させるプログラム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法、およびプログラムに関する。

【背景技術】

【0002】

システム運用オペレーションで利用する各種データを AI に学習させて、オペレーションの自動化、効率化を図る Artificial Intelligence for IT Operations (AIOps) が知られている。システム運用オペレーションは判断の説明責任を伴うが、AI はモデルがブラックボックスとなる場合があるため、AIOps による判断の説明情報を確保する必要がある。

30

【0003】

例えば、判断構造の概観（大局的説明情報）は、ベテランオペレータの臨機応変な判断を模擬したベイジアンネットワークで示すことができる。ベイジアンネットワークは、ノード（判断要素）、エッジ（判断要素同士の関係）、および条件付き確率表（判断要素の影響度合い）によるグラフィカルモデルにより、人の経験（ドメイン知識）を反映した判断モデルを設計可能である。ベイジアンネットワークにより、観測済み情報を元に未観測の情報を確率計算で推論し、AI 出力の妥当性を検証できる。また、AI 個別の判断根拠の詳細（局所的説明情報）は、SHapley Additive exPlanations (SHAP) を用いて生成できる。

40

【先行技術文献】

【非特許文献】

【0004】

【文献】HUGIN EXPERT, “Building a Bayesian Network”, < URL : <https://hugin.com/wp-content/uploads/2016/05/Building-a-BN-Tutorial.pdf> >

【発明の概要】

【発明が解決しようとする課題】

【0005】

ベイジアンネットワークに入力するデータは、ワークフロー自動化の一部としてシステ

50

ムから入力されることが多い。しかしながら、インプットのソースとするシステムのデータ自体が手動入力の場合があり、ヒューマンエラーが混在し、ベイジアンネットワークに誤ったデータを入力するリスクがある。誤ったデータを発見して修正を促すことで、ベイジアンネットワークの周辺環境をメンテナンスすることが必要である。

【0006】

データの誤りを検出する手法として、例えば、パリティビット、チェックサムがある。ベイジアンネットワークのノードは離散値（0，1，2，・・・）の組み合わせで表現されるが、ベイジアンネットワーク個体ごとに論理的に意味のある値であるため、パリティビットのように機械的に整合性ルールを定義できない。

【0007】

データ群のなかから外れ値を検出する手法としてLocal Outlier Factor（LOF）があるが、一般的にベイジアンネットワークのノードの値域は0から10程度と狭く、外れ値とみなせるほど大きな差分とならない。

【0008】

本発明は、上記に鑑みてなされたものであり、データの誤りを検出することを目的とする。

【課題を解決するための手段】

【0009】

本発明の一態様の情報処理装置は、ベイジアンネットワークにおけるデータの誤りを検出する情報処理装置であって、投入したデータに基づくベイジアンネットワークの各ノードの判断の傾向についてケンドールの一致度係数を計算する計算部と、前記ケンドールの一致度係数が閾値より低い場合に、前記データが誤りを含む旨の判定結果を出力する出力部を備える。

【0010】

本発明の一態様の情報処理方法は、ベイジアンネットワークにおけるデータの誤りを検出する情報処理方法であって、コンピュータが、投入したデータに基づくベイジアンネットワークの各ノードの判断の傾向についてケンドールの一致度係数を計算し、前記ケンドールの一致度係数が閾値より低い場合に、前記データが誤りを含む旨の判定結果を出力する。

【発明の効果】

【0011】

本発明によれば、データの誤りを検出できる。

【図面の簡単な説明】

【0012】

【図1】図1は、ベイジアンネットワークの一例を示す図である。

【図2】図2は、本実施形態の情報処理装置の構成の一例を示す機能ブロック図である。

【図3】図3は、ベイジアンネットワークの一例を示す図である。

【図4】図4は、各親ノードの子ノードの事後確率値のランキングの一例を示す図である。

【図5】図5は、本実施形態の情報処理装置の処理の流れの一例を示すフローチャートである。

【図6】図6は、情報処理装置のハードウェア構成の一例を示す図である。

【発明を実施するための形態】

【0013】

以下、本発明の実施の形態について図面を用いて説明する。

【0014】

図1を参照し、ベイジアンネットワークについて簡単に説明する。図1のベイジアンネットワークは、がんの診断に関連するベイジアンネットワークの一例である。図1のベイジアンネットワークは、5のノードN1～N5、4つのエッジE1～E4、および各ノードN1～N5の条件付き確率表（CPT）を有する。ノードは判断要素を示し、エッジは判断要素間の因果関係を示す。エッジの矢印の元が親ノードであり、矢印の先が子ノード

10

20

30

40

50

である。ノード間の因果関係はベテランオペレータの知見によって作成できる。図1の例では、ノードN1、N2は、ノードN3の親ノードである。ノードN3は、ノードN4、N5の親ノードである。CPTは、判断要素間の因果関係の度合いを示す。CPTは、例えば、データの統計情報に基づいて人手で算出される。観測済み情報（データ）を親ノードN1、N2に入力すると、未観測のノードN3、N4、N5のノードの確率値が得られる。図1の例では、ノードN3のCancerは、ノードN1のPollutionとノードN2のSmokerの状態（値）によって推論できる。例えば、図1のノードN3のCPTでは、Pollution=high、Smoker=Trueの場合、Cancerである確率値は0.05であり、Pollution=low、Smoker=Falseの場合、Cancerである確率値は0.001である。

10

【0015】

ベテランオペレータの臨機応変な判断を模擬したベイジアンネットワークにより、AIの判断を検証し判断理由を説明できるようにすることで、ネットワークオペレーションにAIを安全に組み込むことができる。

【0016】

次に、図2を参照し、本実施形態の情報処理装置1の構成の一例について説明する。図2に示す情報処理装置1は、入力部11、計算部12、および出力部13を備える。

【0017】

入力部11は、観測済み情報を投入した親ノード群の判断の傾向を求めるための情報を入力する。例えば、入力部11は、親ノードに投入した観測済み情報に基づく、子ノードの、親ノードに対する事後確率値を入力する。事後確率値は、親ノードに投入した観測済み情報とベイジアンネットワークで計算された子ノードの未観測の情報から計算できる。

20

【0018】

図3のベイジアンネットワークの例では、ノードN10、N20、N30にシステムから観測済み情報が投入される。ノードN10、N20、N30に投入された観測済み情報に基づいてノードN40の確率値が得られる。図3の例では、ノードN10に観測済み情報を投入するシステムのデータは、人手によってシステムに投入されるデータであり、誤りが混入するおそれがあるものとする。

【0019】

入力部11は、ノードN10、N20、N30のそれぞれについて、投入した観測済み情報に基づく、子ノードN40の事後確率値を入力する。入力部11は、判断要素がStainのノードN10について、Stain=0である場合のCancer=0となる事後確率値 $P(Cancer=0 | Stain=0)$ とStain=0である場合のCancer=1となる事後確率値 $P(Cancer=1 | Stain=0)$ を入力する。入力部11は、判断要素がPollutionのノードN20について、Pollution=1である場合のCancer=0となる事後確率値 $P(Cancer=0 | Pollution=1)$ とPollution=1である場合のCancer=1となる事後確率値 $P(Cancer=1 | Pollution=1)$ を入力する。入力部11は、判断要素がSmokerのノードN30について、Smoker=1である場合のCancer=0となる事後確率値 $P(Cancer=0 | Smoker=1)$ とSmoker=1である場合のCancer=1となる事後確率値 $P(Cancer=1 | Smoker=1)$ を入力する。親ノードの値の組み合わせは、ワークフロー実行ごとに毎回異なる値がシステムから投入される。上記では、Stain=0、Pollution=1、Smoker=1がベイジアンネットワークに投入されて計算した一例を示した。

30

40

【0020】

計算部12は、親ノード群の判断の傾向についてケンドールの一致度係数を計算することで、観測済み情報同士の整合性を数値化する。具体的には、計算部12は、親ノードごとに、子ノードの事後確率値のランキングを求め、親ノード間でのランキングを用いてケンドールの一致度係数を計算し、観測済み情報同士の整合性を数値化する。

【0021】

50

図4に、図3の親ノードN10, N20, N30のそれぞれについて、子ノードの事後確率値のランキングを求めた一例を示す。子ノードの、親ノードに対する事後確率値は次式の関係であったとする。

【0022】

【数1】

$$\begin{aligned} P(Cancer=0|Stain=0) &> P(Cancer=1|Stain=0) \\ P(Cancer=1|Pollution=1) &> P(Cancer=0|Pollution=1) \\ P(Cancer=1|Smoker=1) &> P(Cancer=0|Smoker=1) \end{aligned}$$

10

【0023】

ノードN10について、事後確率値 $P(Cancer=0|Stain=0)$ が事後確率値 $P(Cancer=1|Stain=0)$ よりも大きかったため、 $Stain=0$ についてのランキングは、 $Cancer=0$ を1位、 $Cancer=1$ を2位とする。

【0024】

ノードN20について、事後確率値 $P(Cancer=0|Pollution=1)$ が事後確率値 $P(Cancer=1|Pollution=1)$ よりも大きかったため、 $Pollution=1$ についてのランキングは、 $Cancer=1$ を1位、 $Cancer=0$ を2位とする。

20

【0025】

ノードN30について、事後確率値 $P(Cancer=0|Smoker=1)$ が事後確率値 $P(Cancer=1|Smoker=1)$ よりも大きかったため、 $Smoker=1$ についてのランキングは、 $Cancer=1$ を1位、 $Cancer=0$ を2位とする。

【0026】

なお、子ノードが3値以上を取る場合は、各親ノードについて3位以下のランキングも求められる。

【0027】

計算部12は、各親ノードについて、子ノードの事後確率値のランキングを求めた後、親ノード間のランキングについてケンドールの一致度係数を求める。ケンドールの一致度係数Wは次式で求められる。

30

【0028】

【数2】

$$\begin{aligned} R_i &= \sum_{j=1}^m r_{ij}, \bar{R} = \frac{1}{n} \sum_{i=1}^n R_i, S = \sum_{i=1}^n (R_i - \bar{R})^2 \\ W &= \frac{12S}{m^2 n (n^2 - 1)} \end{aligned}$$

40

【0029】

ここで、 i は子ノードの各値（例えば $Cancer=0$, $Cancer=1$ ）、 j は親ノード（例えばノードN10, N20, N30）、 r_{ij} は親ノード j による子ノードの値 i のランキング値（例えば1位または2位）、 n は子ノードの値の数、 m は親ノードの数、 R_i は子ノードの値 i ごとの順位の和、 $\bar{R}$ （上にバー）は順位の和の平均、 S は順位に関する平方和である。

【0030】

親ノード間の判断が整合していればケンドールの一致度係数Wは1に近づき、整合していなければケンドールの一致度係数Wは0に近づく。

【0031】

50

出力部 13 は、計算部 12 の求めた整合性の数値（ケンドールの一致度係数 W ）が任意の閾値よりも低い場合に、投入された観測済み情報が誤りを含む可能性があることを示す判定結果を出力し、修正を促す。例えば、出力部 13 は、親ノードの子ノードに与える作用の方向を事後確率値から評価し、判断の傾向が他と異なるノードに投入された観測済み情報の誤りの可能性が高いとみなして修正を促す。

【0032】

次に、図 5 のフローチャートを参照し、本実施形態の情報処理装置 1 の処理の流れの一例について説明する。

【0033】

ステップ S1 にて、情報処理装置 1 は、親ノード群の判断の傾向を求める。具体的には、情報処理装置 1 は、各親ノードについて子ノードの事後確率値を入力し、親ノードごとに子ノードの事後確率値のランキングを求める。

10

【0034】

ステップ S2 にて、情報処理装置 1 は、親ノード群の判断の傾向についてケンドールの一致度係数を計算する。具体的には、情報処理装置 1 は、ステップ S1 で求めたランキングの一致度についてケンドールの一致度係数を計算する。

【0035】

ステップ S3 にて、情報処理装置 1 は、誤ったデータが混入した可能性を判定する。具体的には、情報処理装置 1 は、ステップ S2 で計算したケンドールの一致度係数と所定の閾値とを比較し、ケンドールの一致度係数が所定の閾値よりも低い場合に、データが誤りを含む可能性があることを示す判定結果を出力する。このとき、情報処理装置 1 は、親ノードの子ノードに与える作用の方向を事後確率から評価し、判断の傾向が他と異なるノードに投入されたデータが誤りを含む可能性があることを示してもよい。

20

【0036】

以上説明したように、本実施形態の情報処理装置 1 は、ベイジアンネットワークのノードに投入した観測済み情報に基づく各ノードの判断の傾向についてケンドールの一致度係数を計算する計算部 12 と、ケンドールの一致度係数が閾値より低い場合に、データが誤りを含む旨の判定結果を出力する出力部 13 を備える。計算部 12 は、観測済み情報に基づく各ノードについて子ノードの事後確率値のランキングを求め、求めたランキングについてケンドールの一致度係数を計算する。これにより、誤ったデータを検出することができ、誤ったデータの修正を促して、ベイジアンネットワークの周辺環境をメンテナンスすることを可能にする。

30

【0037】

上記説明した情報処理装置 1 には、例えば、図 6 に示すような、中央演算処理装置（CPU）901 と、メモリ 902 と、ストレージ 903 と、通信装置 904 と、入力装置 905 と、出力装置 906 とを備える汎用的なコンピュータシステムを用いることができる。このコンピュータシステムにおいて、CPU 901 がメモリ 902 上にロードされた所定のプログラムを実行することにより、情報処理装置 1 が実現される。このプログラムは磁気ディスク、光ディスク、半導体メモリなどのコンピュータ読み取り可能な記録媒体に記録することも、ネットワークを介して配信することもできる。

40

【符号の説明】

【0038】

1 情報処理装置

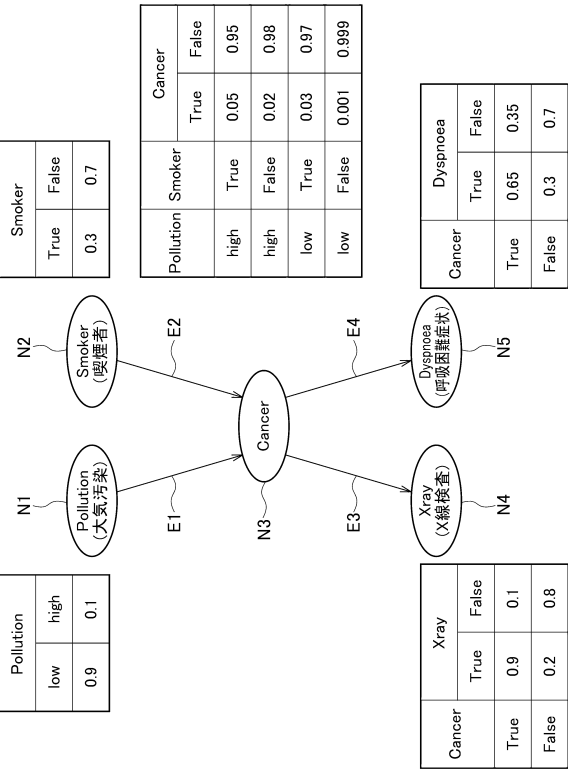
11 入力部

12 計算部

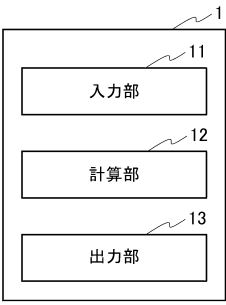
13 出力部

【図面】

【図 1】



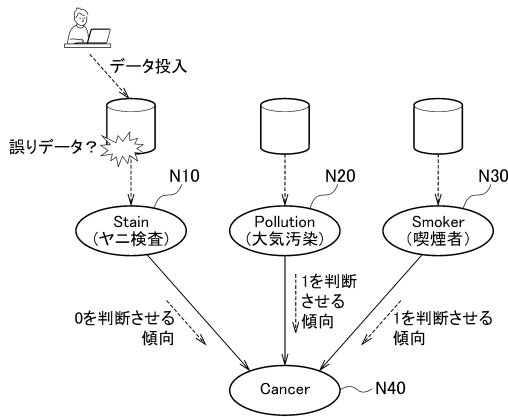
【図 2】



10

20

【図 3】



【図 4】

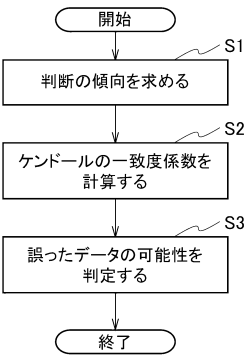
Cancerの インデックス	Stain=0	Pollution=1	Smoker=1
0	1	2	2
1	2	1	1

30

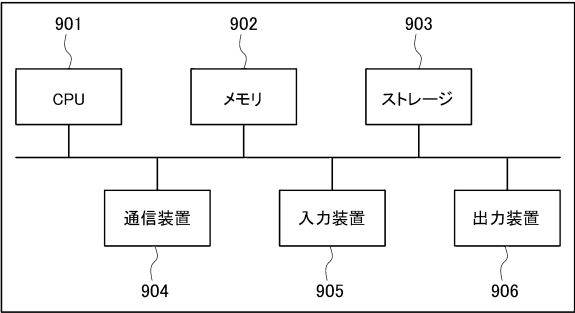
40

50

【図 5】



【図 6】



10

20

30

40

50

フロントページの続き

日本電信電話株式会社内

(72)発明者 中村 瑞人

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

審査官 福西 章人

(56)参考文献 特開2018-124829 (JP, A)

特表2012-504810 (JP, A)

特開2014-85948 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06N 3/00-99/00

G06F 18/00-18/40