

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第5854274号
(P5854274)

(45) 発行日 平成28年2月9日 (2016.2.9)

(24) 登録日 平成27年12月18日 (2015.12.18)

(51) Int.Cl.

F I

GO6F 17/30 (2006.01)

GO6N 5/04 (2006.01)

GO6F 17/30 419B

GO6F 17/30 220Z

GO6N 5/04 550J

請求項の数 13 (全 22 頁)

(21) 出願番号	特願2012-72789 (P2012-72789)	(73) 特許権者	000002185
(22) 出願日	平成24年3月28日 (2012.3.28)		ソニー株式会社
(65) 公開番号	特開2013-206016 (P2013-206016A)		東京都港区港南1丁目7番1号
(43) 公開日	平成25年10月7日 (2013.10.7)	(74) 代理人	100082131
審査請求日	平成26年12月26日 (2014.12.26)		弁理士 稲本 義雄
		(74) 代理人	100121131
			弁理士 西川 孝
		(72) 発明者	磯崎 隆司
			東京都品川区東五反田3丁目14番13号
			株式会社ソニーコンピュータサイエンス
			研究所内
		審査官	山本 俊介

最終頁に続く

(54) 【発明の名称】 情報処理装置および方法、並びにプログラム

(57) 【特許請求の範囲】

【請求項 1】

多変数間の独立性を検定する情報処理装置において、
グラフィカルモデルにおいて少なくとも独立ではない2変数間の経路上に、前記2変数間の独立性の条件となる条件変数がある場合、前記2変数間の条件付き独立性検定を実行する実行部と、

独立である第1および第2の変数がそれぞれ、第3の変数に対して独立でないグラフ構造であるV字構造を探索するV字構造探索部と、

探索された前記V字構造を記憶するV字構造記憶部と、

前記2変数の経路上に、前記V字構造記憶部に記憶されている前記V字構造と一致するグラフ構造があるか否かによって、前記V字構造の有無を判定する判定部と

を備え、

前記実行部は、前記V字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記2変数間の条件付き独立性検定を実行しない
情報処理装置。

【請求項 2】

前記実行部は、前記条件変数が、前記V字構造における前記第3の変数である場合には、前記2変数間の条件付き独立性検定を実行しない

請求項1に記載の情報処理装置。

【請求項 3】

前記実行部は、条件付き独立性検定を実行するより前に、前記第 1 および第 2 の変数間の条件のない独立性検定を実行し、

前記 V 字構造探索部は、前記第 1 および第 2 の変数が条件のない独立であり、かつ、前記第 1 および第 2 の変数がそれぞれ前記第 3 の変数に対して独立でないグラフ構造を前記 V 字構造として探索する

請求項 1 または 2 に記載の情報処理装置。

【請求項 4】

前記 V 字構造探索部は、前記第 1 および第 2 の変数が、前記第 3 の変数以外の変数を条件として条件付き独立であり、かつ、前記第 1 および第 2 の変数がそれぞれ前記第 3 の変数に対して独立でないグラフ構造を前記 V 字構造として探索する

10

請求項 3 に記載の情報処理装置。

【請求項 5】

前記 V 字構造探索部は、インクリメントされる前記条件変数の組の数が所定の数を超えた場合、前記 V 字構造を探索しない

請求項 4 に記載の情報処理装置。

【請求項 6】

前記判定部は、インクリメントされる前記条件変数の組の数に応じた数の前記 2 変数間の経路について、前記 V 字構造の有無を判定し、

前記実行部は、前記 V 字構造を有すると判定された経路のうちのいずれかに、前記条件変数のいずれかがある場合には、前記 2 変数間の条件付き独立性検定を実行しない

20

請求項 1 乃至 5 のいずれかに記載の情報処理装置。

【請求項 7】

複数の前記 V 字構造において、矛盾する有向辺を検出し、前記有向辺の一部または全部を無向辺にすることで、前記矛盾を解消する矛盾検出部をさらに備える

請求項 1 乃至 6 のいずれかに記載の情報処理装置。

【請求項 8】

前記矛盾検出部は、複数の前記 V 字構造における前記第 1 および第 2 の変数間の独立性検定に用いられる指標を用いて、より独立性の弱い前記第 1 および第 2 の変数を有する前記 V 字構造の有向辺を無向辺にする

請求項 7 に記載の情報処理装置。

30

【請求項 9】

変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない第 1 の V 字構造、および、変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない第 2 の V 字構造において、変数 X と変数 W とが独立でない場合、前記矛盾検出部は、前記第 1 の V 字構造における変数 X から変数 Z への有向辺、および、前記第 2 の V 字構造における変数 W から変数 Z への有向辺を無向辺にする

請求項 7 に記載の情報処理装置。

【請求項 10】

変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない第 1 の V 字構造、および、変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない第 2 の V 字構造において、変数 X と変数 W とが独立でない場合、前記矛盾検出部は、前記第 1 および第 2 の V 字構造における有向辺の全部を無向辺にする

40

請求項 7 に記載の情報処理装置。

【請求項 11】

変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない第 1 の V 字構造、および、変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない第 2 の V 字構造において、変数 X と変数 W とが独立でない場合、前記矛盾検出部は、2 変数間の独立性検定に用いられる指標を用いて、変数 X および変数 Y 間の独立性と、変数 W および変数 Y 間の独立性とで、より独立性の弱い 2 変数を有する前記 V 字構造の有向辺を無向辺にする

請求項 7 に記載の情報処理装置。

50

【請求項 1 2】

多変数間の独立性を検定する情報処理装置の情報処理方法において、

グラフィカルモデルにおいて少なくとも独立ではない2変数間の経路上に、前記2変数間の独立性の条件となる条件変数がある場合、前記2変数間の条件付き独立性検定を実行する実行ステップと、

独立である第1および第2の変数がそれぞれ、第3の変数に対して独立でないグラフ構造であるV字構造を探索するV字構造探索ステップと、

探索された前記V字構造を記憶するV字構造記憶ステップと、

前記2変数の経路上に、前記V字構造記憶ステップにより記憶された前記V字構造と一致するグラフ構造があるか否かによって、前記V字構造の有無を判定する判定ステップと

10

を含み、

前記実行ステップは、前記V字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記2変数間の条件付き独立性検定を実行しない

情報処理方法。

【請求項 1 3】

多変数間の独立性を検定する処理をコンピュータに実行させるプログラムにおいて、

グラフィカルモデルにおいて少なくとも独立ではない2変数間の経路上に、前記2変数間の独立性の条件となる条件変数がある場合、前記2変数間の条件付き独立性検定を実行する実行ステップと、

独立である第1および第2の変数がそれぞれ、第3の変数に対して独立でないグラフ構造であるV字構造を探索するV字構造探索ステップと、

20

探索された前記V字構造を記憶するV字構造記憶ステップと、

前記2変数の経路上に、前記V字構造記憶ステップにより記憶された前記V字構造と一致するグラフ構造があるか否かによって、前記V字構造の有無を判定する判定ステップと

を含む処理をコンピュータに実行させ、

前記実行ステップは、前記V字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記2変数間の条件付き独立性検定を実行しない

プログラム。

【発明の詳細な説明】

【技術分野】

30

【0001】

本技術は、情報処理装置および方法、並びにプログラムに関し、特に、多変数間の因果関係の推定結果の信頼性をより高めることができるようにする情報処理装置および方法、並びにプログラムに関する。

【背景技術】

【0002】

従来、多変量確率変数における観察データからの統計的因果関係の推定は、情報量規準や罰則付き最尤法またはベイズ法による推定結果をスコアとしてこれを最大化する方法と、変数間の条件付き独立性の統計的検定によって推定する方法とに大別される。その結果としての変数間の因果関係を、グラフィカルモデル（非循環モデル）として表現することは、その結果の可読性の良さからしばしば行われている。

40

【0003】

図1は、変数Xと変数Yとの因果関係を表す3つのグラフィカルモデルの例を示している。

【0004】

図1上段に示されるグラフィカルモデルは、変数Xと変数Yとの因果関係が不明な場合のモデルであり、変数Xと変数Yとを頂点として、それぞれが、方向のない辺（無向辺）で結ばれている。また、図1中段に示されるグラフィカルモデルは、変数Xと変数Yとの因果関係として、変数Xが原因、変数Yが結果である場合のモデルであり、変数Xと変数Yとを頂点として、それぞれが、原因から結果への方

50

る。そして、図 1 下段に示されるグラフィカルモデルは、変数 X と変数 Y とを頂点として、それぞれが、3 つの変数とそれぞれを結ぶ辺で結ばれている。図 1 下段に示されるグラフィカルモデルにおいて、3 つの変数とそれぞれを結ぶ辺は、変数 X と変数 Y との間の経路であり、その一部に、方向のある有向辺を含むようにできる。

【 0 0 0 5 】

さて、変数間の条件付き独立性の統計的検定によって変数間の因果関係を推定する方法は、潜在的な共通原因変数の存在を推測できる可能性があることと、有向辺の方向付けの理由に物理学的な背景を有することから、重要な方法とされている。しかしながら、一方では、統計的検定の検出力不足のため、推定結果の信頼性が低いことが長年にわたり解決されていない。

10

【 0 0 0 6 】

この方法における初期の試みでは、 n 変数確率変数における 2 変数間の条件付き独立性検定を行うには、その条件となる条件変数の組として、 $(n-2)$ 個の変数から、最小で 0 個、最大で $(n-2)$ 個の変数の組み合わせを全て取り出して、総当たりで検定を行うことが考えられた。しかしながら、この場合、変数の組み合わせ数が指数関数的に増大するため、計算機による計算は現実的ではなかった。

【 0 0 0 7 】

その後、非循環有向グラフにより、変数間の条件付き独立性が一意に表現されると仮定すると、ある 2 変数を条件付き独立にする条件変数の組は一意に定まることと、条件変数は着目する変数と条件付き独立ではないという条件の下で、条件変数の組の数を昇順に増加させて検定を行い、独立性が棄却されない場合には直ちに 2 変数間の辺をはずすことで、検定に要する計算量を大幅に削減するアルゴリズムが開示されている（非特許文献 1 参照）。しかしながら、このアルゴリズムにおいても、検定の試行回数が依然として多いために、検定エラーが頻発してしまう。

20

【 0 0 0 8 】

これに対して、非特許文献 1 には、上述のアルゴリズムを改良し、条件変数は着目する 2 変数間の経路上にあるという条件により、さらに計算量を削減することも開示されている。しかしながら、アルゴリズムの実行途中では、のちに独立と判定されるはずの辺が残っているために長い経路を辿ることが許され、検定の試行回数が実効的には大きく削減されなかった。

30

【 0 0 0 9 】

また、グラフ全体を小さなサブグラフに再帰的に分割することで、大きい組の数の条件変数による独立性検定の試行回数を削減するようにしたアルゴリズムがある（非特許文献 2 参照）。しかしながら、このアルゴリズムでは、小さい組の数の条件変数による独立性検定において、本来独立でないはずが独立であると検定される検定エラーの発生を抑えることはできなかった。

【 0 0 1 0 】

さらに、非特許文献 3 には、非特許文献 2 とは異なる方法で再帰的な処理を行うことが開示されているが、この方法は、独立性検定に大きい数の条件変数を必要とするため、計算の安定性に欠けていた。

40

【 先行技術文献 】

【 非特許文献 】

【 0 0 1 1 】

【 非特許文献 1 】 P. Spirtes, C. Glymour, R. Scheines, "Causation, Prediction, and Search", MIT Press, second edition, 2000.

【 非特許文献 2 】 R. Yehezkel, B. Lerner, "Bayesian Network Structure Learning by Recursive Autonomy Identification", Journal of Machine Learning Research, Vol. 10, pp. 1527-1570, 2009.

【 非特許文献 3 】 X. Xie, Z. Geng, "A Recursive Method for Structural Learning of Directed Acyclic Graphs", Journal of Machine Learning Research, Vol. 9, pp. 459-

50

483, 2008.

【発明の概要】

【発明が解決しようとする課題】

【0012】

このように、条件付き独立性検定による多変数間の因果関係の推定は信頼性が低いことに対して、十分な改善を図る方法はなかった。

【0013】

本技術は、このような状況に鑑みてなされたものであり、多変数間の因果関係の推定結果の信頼性をより高めることができるようにするものである。

【課題を解決するための手段】

【0014】

本技術の一側面の情報処理装置は、多変数間の独立性を検定する情報処理装置であって、グラフィカルモデルにおいて少なくとも独立ではない2変数間の経路上に、前記2変数間の独立性の条件となる条件変数がある場合、前記2変数間の条件付き独立性検定を実行する実行部と、独立である第1および第2の変数がそれぞれ、第3の変数に対して独立でないグラフ構造であるV字構造を探索するV字構造探索部と、探索された前記V字構造を記憶するV字構造記憶部と、前記2変数の経路上に、前記V字構造記憶部に記憶されている前記V字構造と一致するグラフ構造があるか否かによって、前記V字構造の有無を判定する判定部とを備え、前記実行部は、前記V字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記2変数間の条件付き独立性検定を実行しない。

【0015】

前記実行部には、前記条件変数が、前記V字構造における前記第3の変数である場合には、前記2変数間の条件付き独立性検定を実行させないようにすることができる。

【0016】

前記実行部には、条件付き独立性検定を実行するより前に、前記第1および第2の変数間の条件のない独立性検定を実行させ、前記V字構造探索部には、前記第1および第2の変数が条件のない独立であり、かつ、前記第1および第2の変数がそれぞれ前記第3の変数に対して独立でないグラフ構造を前記V字構造として探索させることができる。

【0017】

前記V字構造探索部には、前記第1および第2の変数が、前記第3の変数以外の変数を条件として条件付き独立であり、かつ、前記第1および第2の変数がそれぞれ前記第3の変数に対して独立でないグラフ構造を前記V字構造として探索させることができる。

【0018】

前記V字構造探索部には、インクリメントされる前記条件変数の組の数が所定の数を越えた場合、前記V字構造を探索させないようにすることができる。

【0019】

前記判定部には、インクリメントされる前記条件変数の組の数に応じた数の前記2変数間の経路について、前記V字構造の有無を判定させ、前記実行部には、前記V字構造を有すると判定された経路のうちのいずれかに、前記条件変数のいずれかがある場合には、前記2変数間の条件付き独立性検定を実行させないようにすることができる。

【0020】

複数の前記V字構造において、矛盾する有向辺を検出し、前記有向辺の一部または全部を無向辺にすることで、前記矛盾を解消する矛盾検出部をさらに設けることができる。

【0021】

前記矛盾検出部には、複数の前記V字構造における前記第1および第2の変数間の独立性検定に用いられる指標を用いて、より独立性の弱い前記第1および第2の変数を有する前記V字構造の有向辺を無向辺にさせることができる。

【0022】

変数Xおよび変数Yがそれぞれ、変数Zに対して独立でない第1のV字構造、および、変数Wおよび変数Yがそれぞれ、変数Zに対して独立でない第2のV字構造において、変

10

20

30

40

50

数 X と変数 W とが独立でない場合、前記矛盾検出部には、前記第 1 の V 字構造における変数 X から変数 Z への有向辺、および、前記第 2 の V 字構造における変数 W から変数 Z への有向辺を無向辺にさせることができる。

【 0 0 2 3 】

変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない第 1 の V 字構造、および、変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない第 2 の V 字構造において、変数 X と変数 W とが独立でない場合、前記矛盾検出部には、前記第 1 および第 2 の V 字構造における有向辺の全部を無向辺にさせることができる。

【 0 0 2 4 】

変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない第 1 の V 字構造、および、変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない第 2 の V 字構造において、変数 X と変数 W とが独立でない場合、前記矛盾検出部には、2 変数間の独立性検定に用いられる指標を用いて、変数 X および変数 Y 間の独立性と、変数 W および変数 Y 間の独立性とで、より独立性の弱い 2 変数を有する前記 V 字構造の有向辺を無向辺にさせることができる。

【 0 0 2 5 】

本技術の一側面の情報処理方法は、多変数間の独立性を検定する情報処理装置の情報処理方法であって、グラフィカルモデルにおいて少なくとも独立ではない 2 変数間の経路上に、前記 2 変数間の独立性の条件となる条件変数がある場合、前記 2 変数間の条件付き独立性検定を実行する実行ステップと、独立である第 1 および第 2 の変数がそれぞれ、第 3 の変数に対して独立でないグラフ構造である V 字構造を探索する V 字構造探索ステップと、探索された前記 V 字構造を記憶する V 字構造記憶ステップと、前記 2 変数の経路上に、前記 V 字構造記憶ステップにより記憶された前記 V 字構造と一致するグラフ構造があるか否かによって、前記 V 字構造の有無を判定する判定ステップとを含み、前記実行ステップは、前記 V 字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記 2 変数間の条件付き独立性検定を実行しない。

【 0 0 2 6 】

本技術の一側面のプログラムは、多変数間の独立性を検定する処理をコンピュータに実行させるプログラムであって、グラフィカルモデルにおいて少なくとも独立ではない 2 変数間の経路上に、前記 2 変数間の独立性の条件となる条件変数がある場合、前記 2 変数間の条件付き独立性検定を実行する実行ステップと、独立である第 1 および第 2 の変数がそれぞれ、第 3 の変数に対して独立でないグラフ構造である V 字構造を探索する V 字構造探索ステップと、探索された前記 V 字構造を記憶する V 字構造記憶ステップと、前記 2 変数の経路上に、前記 V 字構造記憶ステップにより記憶された前記 V 字構造と一致するグラフ構造があるか否かによって、前記 V 字構造の有無を判定する判定ステップとを含む処理をコンピュータに実行させ、前記実行ステップは、前記 V 字構造を有すると判定された経路上にのみ、前記条件変数がある場合には、前記 2 変数間の条件付き独立性検定を実行しない。

【 0 0 2 7 】

本技術の一側面においては、グラフィカルモデルにおいて少なくとも独立ではない 2 変数間の経路上に、2 変数間の独立性の条件となる条件変数がある場合、2 変数間の条件付き独立性検定が実行され、独立である第 1 および第 2 の変数がそれぞれ、第 3 の変数に対して独立でないグラフ構造である V 字構造が探索され、探索された V 字構造が記憶され、2 変数の経路上に、記憶された V 字構造と一致するグラフ構造があるか否かによって、V 字構造の有無が判定され、V 字構造を有すると判定された経路上にのみ、条件変数がある場合には、2 変数間の条件付き独立性検定が実行されない。

【 発明の効果 】

【 0 0 2 8 】

本技術の一側面によれば、多変数間の因果関係の推定結果の信頼性をより高めることが可能となる。

10

20

30

40

50

【図面の簡単な説明】**【 0 0 2 9 】****【図 1】** グラフィカルモデルの例を示す図である。**【図 2】** 本技術を適用した情報処理装置のハードウェア構成例を示すブロック図である。**【図 3】** 本技術を適用した情報処理装置の機能構成例を示すブロック図である。**【図 4】** V字構造について説明する図である。**【図 5】** 独立性検定処理について説明するフローチャートである。**【図 6】** 独立性検定処理について説明するフローチャートである。**【図 7】** V字構造の探索について説明する図である。**【図 8】** 2 変数間の経路について説明する図である。

10

【図 9】 2 変数間の経路について説明する図である。**【図 10】** 矛盾解消処理について説明するフローチャートである。**【図 11】** 矛盾解消処理の具体例について説明する図である。**【図 12】** 矛盾解消処理の具体例について説明する図である。**【図 13】** 矛盾解消処理の具体例について説明する図である。**【図 14】** 矛盾解消処理の具体例について説明する図である。**【発明を実施するための形態】****【 0 0 3 0 】**

以下、本技術の実施の形態について図を参照して説明する。

【 0 0 3 1 】

20

[情報処理装置のハードウェア構成例]

図 2 は、本技術を適用した情報処理装置 1 1 のハードウェア構成例を示している。

【 0 0 3 2 】

情報処理装置 1 1 は、離散多変数確率変数間の独立性または条件付き独立性を検定し、その結果をグラフィカルモデルの 1 つであるベイジアンネットワークとして出力する。

【 0 0 3 3 】

情報処理装置 1 1 は、例えば、パーソナルコンピュータで構成することができ、その構成は、パーソナルコンピュータと同様な構成とすることができる。

【 0 0 3 4 】

情報処理装置 1 1 は、CPU (Central Processing Unit) 2 1、ROM (Read Only Memory) 2 2、RAM (Random Access Memory) 2 3、バス 2 4、入出力インタフェース 2 5、入力部 2 6、出力部 2 7、記憶部 2 8、通信部 2 9、およびドライブ 3 0 を含む構成とされている。

30

【 0 0 3 5 】

情報処理装置 1 1 において、CPU 2 1、ROM 2 2、および RAM 2 3 は、バス 2 4 により相互に接続されている。バス 2 4 には、さらに、入出力インタフェース 2 5 が接続されている。入出力インタフェース 2 5 には、キーボード、マウス、タッチパネルなどよりなる入力部 2 6、ディスプレイ、スピーカなどよりなる出力部 2 7、ハードディスクや不揮発性のメモリなどよりなる記憶部 2 8、ネットワークインタフェースなどよりなる通信部 2 9 が接続されている。

40

【 0 0 3 6 】

入出力インタフェース 2 5 にはまた、必要に応じてドライブ 3 0 が接続され、磁気ディスク、光ディスク、光磁気ディスク、あるいは半導体メモリなどよりなるリムーバブルメディア 3 1 が適宜装着され、それらから読み出されたプログラムが、必要に応じて記憶部 2 8 にインストールされる。

【 0 0 3 7 】

また、プログラムは、有線または無線の伝送媒体を介して、通信部 2 9 で受信し、記憶部 2 8 にインストールすることができる。その他、プログラムは、ROM 2 2 や記憶部 2 8 に、あらかじめインストールしておくことができる。

【 0 0 3 8 】

50

なお、情報処理装置 11 が実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われる処理であっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

【0039】

[情報処理装置の機能構成例]

図 3 は、情報処理装置 11 の本技術に係る部分の機能構成例を示している。

【0040】

図 3 の情報処理装置 11 は、入力部 51、制御部 52、記憶部 53、および出力部 54 から構成される。

【0041】

入力部 51 は、図 2 の入力部 26 に対応し、N 個の離散確率変数のうちの、独立性検定の対象となる 2 変数を指定するための引数や、条件付き独立性の条件となる条件変数の変数セット等の入力を受け付け、その内容に対応する情報を制御部 52 に供給する。

【0042】

制御部 52 は、図 2 の CPU 21 に対応し、記憶部 53 に格納されているプログラムに従って動作し、図 2 の記憶部 28 に対応する記憶部 53 に記憶されている各種の情報をを用いて、着目している 2 変数間の独立性検定を実行する。

【0043】

出力部 54 は、図 2 の出力部 27 に対応し、制御部 52 の制御の下、独立性検定の検定結果を、グラフィカルモデルとして出力する。

【0044】

また、図 3 において、制御部 52 は、経路探索部 71、実行判定部 72、検定実行部 73、V 字構造探索部 74、および矛盾検出部 75 を備え、記憶部 53 は、変数対記憶部 81、V 字構造記憶部 82、分離変数組記憶部 83 を備えている。

【0045】

経路探索部 71 は、着目している 2 変数間の経路を、例えば、深さ優先探索等の一般的なアルゴリズムを用いて探索する。

【0046】

実行判定部 72 は、着目している 2 変数について、独立性検定を実行する必要があるか否かを判定する。

【0047】

検定実行部 73 は、着目している 2 変数について、独立性検定を実行する。ただし、実行判定部 72 により独立性検定を実行する必要がないと判定された場合、検定実行部 73 は、独立性検定を実行しない。

【0048】

V 字構造探索部 74 は、独立であると検定された 2 変数について、V 字構造を探索する。V 字構造は、図 4 に示されるように、例えば、ある変数 X と変数 Y とが独立であり、かつ変数 X と変数 Z、変数 Y と変数 Z がそれぞれ独立でないグラフ構造である。

【0049】

矛盾検出部 75 は、V 字構造探索部 74 により探索された V 字構造から、互いに矛盾する関係を有する V 字構造を検出し、その矛盾を解消する。

【0050】

変数対記憶部 81 には、独立でないか、または独立性の検定が行われていない 2 変数が辺で結ばれた変数対が記憶されている。

【0051】

V 字構造記憶部 82 には、V 字構造探索部 74 によって探索された V 字構造が記憶される。なお、実行判定部 72 は、着目している 2 変数について、その 2 変数間の経路上に、V 字構造記憶部 82 に記憶されている V 字構造と一致するグラフ構造があるか否かによって、独立性検定を実行する必要があるか否かを判定する。

【0052】

10

20

30

40

50

分離変数組記憶部 8 3 には、検定実行部 7 3 によって実行される独立性検定により、条件付き独立であると検定された 2 変数間の経路上にある条件変数が、その 2 変数を分離する分離変数組として記憶される。

【 0 0 5 3 】

[情報処理装置による独立性検定処理]

次に、図 5 および図 6 のフローチャートを参照して、情報処理装置 1 1 による独立性検定処理について説明する。記憶部 5 3 には、確率変数（以下、単に変数ともいう）の数 N と、それぞれの変数のとりうる 2 値以上の内部状態数が予め記憶されており、この変数全てについての状態が記述された M 個のデータが、入力部 5 1 により入力されると、独立性検定処理が開始される。

10

【 0 0 5 4 】

なお、初期状態として、変数対記憶部 8 1 には、 N 個の変数がそれぞれ異なる $(N-1)$ 個の変数と対になった（辺で結ばれた）変数対が記憶されており、V 字構造記憶部 8 2 および分離変数組記憶部 8 3 には何も記憶されていない。

【 0 0 5 5 】

ステップ S 1 1 において、制御部 5 2 は、条件変数（以下、条件付き変数ともいう）の組の数 i を 0、すなわち、条件付き変数の組を空集合とする。

【 0 0 5 6 】

ステップ S 1 2 において、制御部 5 2 は、変数対記憶部 8 1 に記憶されている変数対から、1 つの変数対を選択する。

20

【 0 0 5 7 】

ステップ S 1 3 において、検定実行部 7 3 は、独立性検定を実行するのに用いられる値である G^2 統計量を計算する。 G^2 統計量は、以下の式（1）で示される。

【 0 0 5 8 】

【 数 1 】

$$G^2 = 2M \sum_{x, y, z} \hat{P}(x, y, z) \log \frac{\hat{P}(x|y, z)}{\hat{P}(x|z)} \quad \dots (1)$$

【 0 0 5 9 】

式（1）において、 M はデータ数を示している。また、 \wedge （ハット）の付された $P(x, y, z)$ （以下、ハット $P(x, y, z)$ などという）、ハット $P(x|y, z)$ 、ハット $P(x|z)$ はそれぞれ、変数 X 、 Y および変数の組 Z についてのある状態 x 、 y 、 z の推定される結合確率、状態 y 、 z を条件とした場合の状態 x の推定される条件付き確率、状態 z を条件とした場合の状態 x の推定される条件付き確率を表す。これらは、 M 個のデータを用いて推定される。なお、ステップ S 1 3 においては、変数の組 Z は空集合であるので、状態 z は考慮されない。

30

【 0 0 6 0 】

ここでは、上述した G^2 統計量と χ^2 分布とを用いて導き出される、独立性検定の指標となる p 値と、所定の閾値 Th （例えば 5 %（0.05））とが比較されることで、選択された変数対の 2 変数の独立性が検定される。

40

【 0 0 6 1 】

すなわち、ステップ S 1 4 において、検定実行部 7 3 は、 p 値が閾値 Th より大きいかなかを判定する。

【 0 0 6 2 】

ステップ S 1 4 において、 p 値が閾値 Th より大きいと判定された場合、処理はステップ S 1 5 に進み、検定実行部 7 3 は、選択された変数対の 2 変数の独立性を採択する。

【 0 0 6 3 】

そして、ステップ S 1 6 において、検定実行部 7 3 は、選択された変数対を、変数対記憶部 8 1 から削除する。

50

【 0 0 6 4 】

一方、ステップ S 1 4 において、p 値が閾値 Th より大きくないと判定された場合、処理はステップ S 1 7 に進み、検定実行部 7 3 は、選択された変数対の 2 変数の独立性を棄却する。

【 0 0 6 5 】

なお、上述した説明では、独立性検定の指標として p 値を用いるものとしたが、以下の式 (2) で示される条件付き相互情報量 MI を、独立性検定の指標として用いるようにしてもよい。

【 0 0 6 6 】

【 数 2 】

$$MI = \frac{G^2}{2M}$$

・・・ (2)

【 0 0 6 7 】

この場合、条件付き相互情報量 MI と所定の閾値 (例えば 0.05) が比較され、条件付き相互情報量 MI が閾値より小さいと判定された場合に、2 変数の独立性が採択されるようになる。

【 0 0 6 8 】

さて、ステップ S 1 6 またはステップ S 1 7 の後、ステップ S 1 8 において、制御部 5 2 は、変数対記憶部 8 1 に記憶されている全ての変数対について、ステップ S 1 2 乃至ステップ S 1 7 の処理、すなわち条件のない独立性検定を実行したか否かを判定する。

【 0 0 6 9 】

ステップ S 1 8 において、全ての変数対について処理されていないと判定された場合、処理はステップ S 1 2 に戻り、新たに選択される変数対について、ステップ S 1 2 乃至ステップ S 1 7 の処理が繰り返される。

【 0 0 7 0 】

一方、ステップ S 1 8 において、全ての変数対について処理されたと判定された場合、処理はステップ S 1 9 に進む。このとき、変数対記憶部 8 1 は、独立性が棄却された (独立でない) 2 変数の変数対のみが記憶された状態となる。

【 0 0 7 1 】

ステップ S 1 9 において、制御部 5 2 は、上述した処理において、独立な 2 変数があつたか否かを判定する。

【 0 0 7 2 】

ステップ S 1 9 において、独立な 2 変数が 1 組でもあつたと判定された場合、処理はステップ S 2 0 に進み、V 字構造探索部 7 4 は、その独立な 2 変数について、V 字構造を探索する。具体的には、V 字構造探索部 7 4 は、図 4 で示されたように、ある 1 組の変数 X , Y が独立であり、それぞれが変数 Z に対して独立ではない場合、つまり、変数 X , Y が、2 つの辺と変数 Z からなる経路 (X - Z - Y) を有し、かつ変数 Z が変数 X , Y の独立性に条件変数として寄与していない場合、このグラフ構造を V 字構造 (以下、X Z Y 等と表す) として、V 字構造記憶部 8 2 に記憶させる。

【 0 0 7 3 】

一方、ステップ S 1 9 において、独立な 2 変数が 1 組もなかったと判定された場合、ステップ S 2 0 はスキップされる。

【 0 0 7 4 】

その後、処理は図 6 のステップ S 2 1 に進み、制御部 5 2 は、条件付き変数 (条件変数) の組の数 i を 1 インクリメントする。すなわち、条件付き変数の組の数が 1 となる。

【 0 0 7 5 】

ステップ S 2 2 において、制御部 5 2 は、変数対記憶部 8 1 に記憶されている変数対、すなわち、独立でない 2 変数の変数対から、1 つの変数対を選択する。ここでは、変数 S

10

20

30

40

50

と変数 T の変数対が選択されたこととする。

【 0 0 7 6 】

ステップ S 2 3 において、制御部 5 2 は、選択された変数対の 2 変数の独立性の条件となりうる条件付き変数を選択する。具体的には、制御部 5 2 は、変数対記憶部 8 1 に記憶されている変数対を参照することで、選択された変数対の 2 変数のうちのいずれか（例えば変数 S ）と対になっている変数を、条件付き変数として 1 つ選択する。ここでは、変数 W が選択されたこととする。

【 0 0 7 7 】

ステップ S 2 4 において、経路探索部 7 1 は、選択された変数対の 2 変数間の経路を探索し、条件付き変数がその経路上にあるか否かを判定する。

10

【 0 0 7 8 】

ステップ S 2 4 において、例えば、変数 W が、変数 S と変数 T との間の経路上にあると判定された場合、処理はステップ S 2 5 に進み、実行判定部 7 2 は、変数 S と変数 T との間の経路上に、V 字構造記憶部 8 2 に記憶されている V 字構造と一致するグラフ構造があるか否かを判定する。

【 0 0 7 9 】

ステップ S 2 5 において、一致する V 字構造があると判定された場合、その経路では、変数 W は変数 S , T の独立性の条件変数になりえないので、ステップ S 2 6 において、経路探索部 7 1 は、別の経路を探索する。

【 0 0 8 0 】

20

ステップ S 2 7 において、別の経路があるか否かが判定され、別の経路が見つかった場合、処理はステップ S 2 4 に戻り、ステップ S 2 4 , S 2 5 の処理が再度行われる。なお、ステップ S 2 4 において、変数 W が、変数 S と変数 T との間の経路上にないと判定された場合、処理はステップ S 2 3 に戻り、別の条件付き変数が選択される。

【 0 0 8 1 】

一方、ステップ S 2 5 において、一致する V 字構造がないと判定された場合、その経路で、変数 W は変数 S , T の独立性の条件変数になりうるとされ、処理はステップ S 2 8 に進む。

【 0 0 8 2 】

なお、ステップ S 2 8 乃至ステップ S 3 1 およびステップ S 3 3 の処理は、上述したステップ S 1 3 乃至ステップ S 1 7 の処理とそれぞれ同一であるので、その説明は省略するが、ここでは、例えば、変数 W を条件とした変数 S , T の条件付き独立性の検定が実行される。変数 W を条件とした変数 S , T の条件付き独立性が採択された場合には、変数 S , T の変数対が、変数対記憶部 8 1 から削除される。

30

【 0 0 8 3 】

そして、ステップ S 3 2 において、検定実行部 7 3 は、条件付き変数（変数 W ）を、分離変数組として分離変数組記憶部 8 3 に記憶させる。

【 0 0 8 4 】

ところで、ステップ S 2 7 において、別の経路があるか否かが判定され、別の経路が見つからなかった場合、すなわち、V 字構造があると判定された経路上にのみ、条件変数がある場合、ステップ S 2 8 乃至ステップ S 3 3 の処理はスキップされる。つまり、この場合、例えば、変数 W を条件とした変数 S , T の条件付き独立性の検定は実行されず、変数 S , T は、変数対のまま（辺で結ばれたまま）となる。

40

【 0 0 8 5 】

ここで、仮に、変数 S , T が条件付き独立であるならば、条件のない独立性検定（ステップ S 1 3 乃至ステップ S 1 7 の処理）において、すでに独立であると判定されているはずである。これは、非特許文献 1 で開示されている d-separation と faithfulness という理論に基づいている。また、この場合、変数 W は変数 S , T の独立性の条件変数になりえない。したがって、変数 S , T は条件付き独立でないことは明らかであるので、変数 S , T の条件付き独立性の検定は実行されない。

50

【0086】

さて、ステップS32、ステップS33、またはステップS27の後、ステップS34において、制御部52は、変数対記憶部81に記憶されている全ての変数対について、ステップS22乃至ステップS33の処理、すなわち条件付き独立性検定を実行したか否かを判定する。

【0087】

ステップS34において、全ての変数対について処理されていないと判定された場合、処理はステップS22に戻り、新たに選択される変数対について、ステップS22乃至ステップS33の処理が繰り返される。

【0088】

一方、ステップS34において、全ての変数対について処理されたと判定された場合、処理はステップS35に進む。このとき、変数対記憶部81は、独立性が棄却された（独立でない）か、または、条件付き独立性検定が実行されなかった2変数の変数対のみが記憶された状態となる。

【0089】

ステップS35において、制御部52は、上述した処理において、独立（条件付き独立）な2変数があつたか否かを判定する。

【0090】

ステップS35において、独立（条件付き独立）な2変数が1組でもあつたと判定された場合、処理はステップS36に進み、V字構造探索部74は、その独立な2変数について、V字構造を探索する。具体的には、V字構造探索部74は、例えば図7で示されるように、ある1組の変数X、Yが変数Wを条件として条件付き独立であり、それぞれが変数Zに対して独立ではない場合、かつ、変数Zが変数X、Yの条件変数（分離変数組）として分離変数組記憶部83に記憶されていない場合、このグラフ構造をV字構造として、V字構造記憶部82に記憶させる。

【0091】

ただし、独立性検定に誤りが生じれば、ここでのV字構造も誤った構造となり、また、後述するように、条件付き変数の組の数が大きくなるほど、検定の誤りも生じやすくなるため、ここで得られたV字構造が検定の対象となる2変数間の経路上にあつたとしても検定を実行させるようにしてもよい。

【0092】

一方、ステップS35において、独立（条件付き独立）な2変数が1組もなかったと判定された場合、ステップS36はスキップされる。

【0093】

ステップS37において、制御部52は、条件付き変数（条件変数）の組の数 i が、最大の数（変数セットに含まれる全ての組の数） i_{\max} （ $i = i_{\max}$ ）であるか否かを判定する。

【0094】

ステップS37において、 $i = i_{\max}$ でないと判定された場合、処理はステップS21に戻り、条件付き変数（条件変数）の組の数 i が1インクリメントされて、これ以降の処理が行われる。

【0095】

例えば、 $i = 2$ となった場合、2個の条件付き変数それぞれについて、 $i = 1$ の場合と同様の処理が行われ、2個の条件付き変数がともに1個の場合と同じ条件を満たす場合のみ、条件付き独立性検定が実行される。すなわち、V字構造がある経路のいずれかに、条件付き変数のいずれかがある場合には、条件付き独立性検定は実行されない。これにより、不要な条件付き独立性検定の実行が回避され、エラーの生じる回数を削減することができる。

【0096】

そして、ステップS37において、 $i = i_{\max}$ であると判定された場合、処理は終了する

10

20

30

40

50

。このようにして、実行可能な全ての検定が終了すると、変数対記憶部 8 1 には、検定により独立ではないと判定され続けたか、または、検定されずに独立ではないとされた 2 変数の変数対が残る。また、V 字構造記憶部 8 2 には、V 字構造となる変数の組が記憶されるようになり、分離変数組記憶部 8 3 には、独立である 2 変数についての条件変数の組が記憶されるようになる。

【 0 0 9 7 】

なお、記憶部 5 3 に記憶されている内容は、制御部 5 2 の制御の下、出力部 5 4 に出力される。具体的には、有向辺と無向辺とが混在した部分有向非循環グラフ（ベイジアンネットワーク）が出力され、変数間の依存関係や因果関係の推定が行われるようになる。

【 0 0 9 8 】

以上の処理によれば、2 変数間の経路のうち、V 字構造を有する経路にのみ、条件変数がある場合には、2 変数間の条件付き独立性検定は実行されなくなる。

【 0 0 9 9 】

例えば、変数 X と変数 Y との間の経路として、図 8 に示される経路 1 のみが探索され、その経路 1 には V 字構造があり、かつ、経路 1 上に条件変数 Z がある場合には、変数 X と変数 Y との間の条件付き独立性検定は実行されない。ただし、変数 X と変数 Y との間の経路として、V 字構造がなく条件付き変数 Z がある経路 2 が探索された場合には、条件付き独立性検定は実行される。

【 0 1 0 0 】

また、変数 X と変数 Y との間の経路として、図 9 に示される経路のみが探索され、その経路には V 字構造があり、かつ、その V 字構造における独立な 2 変数それぞれに対して独立でない変数が条件変数 Z である場合にも、変数 X と変数 Y との間の条件付き独立性検定は実行されない。

【 0 1 0 1 】

このように、以上の処理によれば、条件付き独立性検定の実行回数を削減することができ、検定エラーの発生の頻度を低くすることができ、ひいては、変数間の因果関係の推定結果の信頼性をより高めることが可能となる。

【 0 1 0 2 】

なお、以上においては、離散値としての変数について独立性検定を実行する処理について説明してきたが、変数が連続値である場合には、条件付き独立性の指標としての偏相関係数と、フィッシャーの Z 変換とを用いることで、独立性検定を実行することができる。また、統計的なパターン認識に用いられる手法であるカーネル法を用いることで、独立性検定を実行するようにしてもよい。

【 0 1 0 3 】

また、上述した独立性検定処理において、条件変数の組の数が大きくなれば、条件付き独立性検定の信頼性の低下は避けられず、したがって、V 字構造の探索の精度も悪化し、結果として、変数間の因果関係の推定結果の信頼性を低下させる恐れがある。

【 0 1 0 4 】

そこで、インクリメントされる条件付き変数（条件変数）の組の数 i が、所定の数の超えた場合には、ステップ S 3 5 において条件付き独立な 2 変数があったと判定されたとしても、ステップ S 3 6 の処理（V 字構造の探索）をスキップするようにしてもよい。

【 0 1 0 5 】

これにより、条件付き独立性検定の信頼性の低下による、精度の悪い V 字構造の探索は行われなくなり、変数間の因果関係の推定結果の信頼性の低下を避けることができる。なお、スキップされた V 字構造の探索は、全ての条件付き独立性検定が終わった後に実行されるものとする。

【 0 1 0 6 】

ところで、上述した独立性検定処理により、V 字構造記憶部 8 2 には、探索された V 字構造が記憶されるようになるが、独立性検定処理は、それぞれの 2 変数間について独立して実行されるため、検定のエラーが生じた場合、互いに矛盾した関係を有する V 字構造が

10

20

30

40

50

記憶される恐れがある。

【 0 1 0 7 】

[矛盾解消処理について]

そこで、図 1 0 のフローチャートを参照して、V 字構造の矛盾を解消する矛盾解消処理について説明する。

【 0 1 0 8 】

ステップ S 5 1 において、矛盾検出部 7 5 は、V 字構造記憶部 8 2 に記憶されている V 字構造の中から、矛盾のある V 字構造を検出する。

【 0 1 0 9 】

ステップ S 5 2 において、矛盾検出部 7 5 は、所定の基準にしたがって、検出した V 字構造の矛盾を解消する。

【 0 1 1 0 】

[矛盾解消処理の具体例 1]

図 1 1 は、互いに矛盾した関係を有する V 字構造に対する矛盾解消処理の具体例について説明する図である。

【 0 1 1 1 】

図 1 1 上側には、独立である変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない V 字構造 (X Z Y) と、独立である変数 W および変数 Z がそれぞれ、変数 X に対して独立でない V 字構造 (W X Z) とが示されている。

【 0 1 1 2 】

この場合、一方では X Z となる関係があり、他方では X Z となる関係があるという矛盾が生じている。

【 0 1 1 3 】

これに対して、矛盾検出部 7 5 は、それぞれの V 字構造において独立である 2 変数についての独立性検定に用いられた条件付き独立性の指標、具体的には、p 値や条件付き相互情報量 MI を用いて、条件付き独立性の弱い 2 変数を有する V 字構造の有向辺を無向辺にする。

【 0 1 1 4 】

具体的には、変数 X および変数 Y、変数 W および変数 Z のそれぞれについての p 値が比較されることで、いずれの 2 変数間の条件付き独立性が弱いかが判定される。例えば、変数 X および変数 Y についての p 値が 0.50 であり、変数 W および変数 Z についての p 値が 0.20 であった場合には、変数 X および変数 Y を有する X Z Y の V 字構造が採用され、図 1 1 下側に示されるように、変数 Z を原因とし、変数 X を結果とした因果関係 (以下、Z X の関係などと表す) を表す有向辺が無向化される。また、V 字構造記憶部 8 2 からは W X Z の V 字構造が削除される。

【 0 1 1 5 】

[矛盾解消処理の具体例 2]

図 1 2 は、互いに矛盾した関係を有する V 字構造に対する矛盾解消処理の他の具体例について説明する図である。

【 0 1 1 6 】

図 1 2 上側には、独立である変数 X および変数 Y がそれぞれ、変数 Z に対して独立でない V 字構造 (X Z Y) と、独立である変数 W および変数 Y がそれぞれ、変数 Z に対して独立でない V 字構造 (W Z Y) とが示されている。また、図 1 2 において、変数 X と変数 W とは独立でない。

【 0 1 1 7 】

この場合、変数 X と変数 W とは、本来は、変数 Z 以外の変数を条件変数として条件付き独立である必要があるという矛盾が生じている。

【 0 1 1 8 】

これに対して、矛盾検出部 7 5 は、図 1 2 下側に示されるように、X Z Y の V 字構造における X Z、および、W Z Y の V 字構造における W Z の関係を表す有向辺を

10

20

30

40

50

無向辺にするとともに、それぞれのV字構造を、V字構造記憶部82から削除する。

【0119】

[矛盾解消処理の具体例3]

図13は、互いに矛盾した関係を有するV字構造に対する矛盾解消処理のさらに他の具体例について説明する図である。

【0120】

図13上側には、図12と同様、独立である変数Xおよび変数Yがそれぞれ、変数Zに対して独立でないV字構造(X Z Y)と、独立である変数Wおよび変数Yがそれぞれ、変数Zに対して独立でないV字構造(W Z Y)とが示されている。また、図13においても、変数Xと変数Wとは独立でない。

10

【0121】

この場合も、変数Xと変数Wとは、本来は、変数Z以外の変数を条件変数として条件付き独立である必要があるという矛盾が生じている。

【0122】

これに対して、矛盾検出部75は、図13下側に示されるように、X Z YのV字構造、および、W Z YのV字構造における全ての有向辺を無向辺にするとともに、それぞれのV字構造を、V字構造記憶部82から削除する。

【0123】

[矛盾解消処理の具体例4]

図14は、互いに矛盾した関係を有するV字構造に対する矛盾解消処理のさらに他の具体例について説明する図である。

20

【0124】

図14上側には、図12と同様、独立である変数Xおよび変数Yがそれぞれ、変数Zに対して独立でないV字構造(X Z Y)と、独立である変数Wおよび変数Yがそれぞれ、変数Zに対して独立でないV字構造(W Z Y)とが示されている。また、図14においても、変数Xと変数Wとは独立でない。

【0125】

この場合も、変数Xと変数Wとは、本来は、変数Z以外の変数を条件変数として条件付き独立である必要があるという矛盾が生じている。

【0126】

これに対して、矛盾検出部75は、それぞれのV字構造において独立である2変数についての独立性検定に用いられた条件付き独立性の指標、具体的には、p値や条件付き相互情報量MIを用いて、条件付き独立性の弱い2変数を有するV字構造において、他方のV字構造に共有されない有向辺を無向辺にする。

30

【0127】

具体的には、変数Xおよび変数Y、変数Wおよび変数Yのそれぞれについてのp値が比較され、いずれの2変数間の条件付き独立性が弱いかが判定される。例えば、変数Xおよび変数Yについてのp値が0.50であり、変数Wおよび変数Yについてのp値が0.30であった場合には、変数Xおよび変数Yを有するX Z YのV字構造が採用され、図14下側に示されるように、W Z YのV字構造において、X Z YのV字構造に共有されないW Zの関係を表す有向辺が無向化される。また、V字構造記憶部82からはW Z YのV字構造が削除される。

40

【0128】

以上の処理によれば、V字構造において、矛盾するV字構造が検出され、その矛盾が解消される。したがって、矛盾解消処理を、独立性検定処理と並行して実行することで、誤ったV字構造を用いた条件付き独立性検定によるエラーの発生する可能性を低減することができ、ひいては、変数間の因果関係の推定結果の信頼性をより高めることが可能となる。

【0129】

[本技術の適用例]

50

なお、本技術は、以下に挙げるような例に適用することが可能である。

【0130】

(1) 統計的因果関係推定装置

多変量確率変数における観察データからの統計的因果関係を推定する因果関係推定処理に、本技術の独立性検定処理を適用する。なお、離散値または連続値である多変数確率変数はユーザによって定義され、データセットも予め用意されているものとする。これにより、有向辺と無向辺の混在したグラフィカルモデルが、出力部54としてのモニタ等に画像として出力されるようになる。このとき、条件付き独立性を考慮して有向辺の向きを定めるオリエンテーション・ルールにしたがって、グラフィカルモデルにおける無向辺を、できる限り有向辺に置き換えるようにしてもよい。また、グラフィカルモデルが、同等の関係を表すテキストデータとして、モニタ等に出力されるようにしてもよい。

10

【0131】

(2) 医療診断支援装置

病気と症状との間の因果関係を推定する因果関係推定処理に、本技術の独立性検定処理を適用する。なお、確率変数として、{結核T, 肺がんL, 気管支炎B, 胸部X線検査結果X, 呼吸困難D, および喫煙状況S}が定義され、それぞれの内部状態数がデータとして用意されているものとする。これにより、V字構造や無向辺を有する変数対が判明し、有向辺と無向辺の混在したグラフィカルモデルが、出力部54としてのモニタ等に画像として出力されるようになる。また、条件付き独立性を考慮して有向辺の向きを定めるオリエンテーション・ルールにしたがって、グラフィカルモデルにおける無向辺を、できる限り有向辺に置き換えるようにしてもよい。この例では、変数間の因果関係として、例えば、{S-L}, {S-B}, {L-D}, {B-D}, {L-X}, {T-X}等が出力される。

20

【0132】

(3) 生産管理装置

工場における様々な測定項目と歩留りとの間の因果関係を推定する因果関係推定処理に、本技術の独立性検定処理を適用する。なお、確率変数として、{良品か否かY, 工場内温度T, 工場内湿度M, 測定項目1A, 測定項目2B, および測定項目3C}が定義されているものとする。ここで、変数Yは、良品であるか、良品でないかの2値を有する離散値であるが、他の変数は全て連続値であるものとする。そこで、離散値である変数Yを基準の変数として、他の変数(連続値)を全て離散化する既知の手法を用いることで、全ての変数を離散化する。これにより、有向辺と無向辺の混在したグラフィカルモデルが、出力部54としてのモニタ等に画像として出力されるようになる。

30

【0133】

(4) 文書分類装置

入力された文書をユーザにより予め定義されたカテゴリに分類する文書分類装置における文書分類の学習処理に、本技術の独立性検定処理を適用する。この文書分類装置は、予め分類された、または分類タグが付された文書が入力されると、その文書を、形態素解析等により品詞に分解する。これで、カテゴリを表す変数Cと、品詞の出現頻度を表す変数Sが用意されたことになる。ここで、変数Cは離散値であるが、変数Sは連続値であるので、変数Cを基準として変数Sを離散化(カテゴリ化)する。そして、離散値である変数Sについて、本技術の独立性検定処理を実行することで、有向辺と無向辺の混在したグラフィカルモデルが得られる。このとき、情報量規準を用いて、無向辺を有向辺に置き換えるようにしてもよい。さらに、全ての変数Sとカテゴリを表す変数Cとを有向辺で結ぶことで、条件付き確率分布の積がベイジアンネットワークとして表現されるようになる。これにより、同じデータを用いて条件付き確率分布を推定することで、文書分類の学習処理が行われる。このようにして、品詞間の依存関係を推定(学習)することで、未知の文書の分類に対して、精度の高い分類器が生成されるようになる。

40

【0134】

この文書分類装置に、分類されていない新しい文書が入力されると品詞に分解され、分

50

解された品詞の頻度情報が、分類器における変数Sの頻度情報カテゴリにマッピングされることで、入力された文書に、品詞それぞれの出現確率を表す変数Sが証拠として与えられる。そして、証拠の与えられていないカテゴリの確率値が、変数Sを周辺化することで計算され、最大となる確率値が計算されたカテゴリに、入力された文書が分類されるようになる。

【0135】

(5) 画像識別装置

入力された画像をユーザにより予め定義されたカテゴリに分類する画像識別装置における画像分類の学習処理に、本技術の独立性検定処理を適用する。この画像識別装置は、予め分類された、または分類タグが付された画像（学習用画像）が複数入力されると、これらの画像をクラスタ化する。さらに、画像識別装置は、クラスタ化された画像それぞれのRGB値の平均値を用いて、これらの画像を教師なしでさらにクラスタ化する。これで、カテゴリを表す変数Cと、RGB値の平均値を表す変数Vが用意されたことになる。ここで、変数Cは離散値であるが、変数Vは連続値であるので、変数Cを基準として変数Vを離散化（カテゴリ化）する。そして、離散値である変数Vについて、本技術の独立性検定処理を実行することで、有向辺と無向辺の混在したグラフィカルモデルが得られる。このとき、情報量規準を用いて、無向辺を有向辺に置き換えるようにしてもよい。さらに、全ての変数Vとカテゴリを表す変数Cとを有向辺で結ぶことで、条件付き確率分布の積がベイジアンネットワークとして表現されるようになる。これにより、同じデータを用いて条件付き確率分布を推定することで、画像分類の学習処理が行われる。このようにして、RGB値間の依存関係を推定（学習）することで、未知の画像の分類に対して、精度の高い分類器が生成されるようになる。

【0136】

この画像識別装置に、分類されていない新しい画像が入力されるとRGB値の平均値が算出され、その画像のRGB値の平均値が、分類器における変数VのRGB値カテゴリにマッピングされることで、入力された画像に、変数Vが証拠として与えられる。そして、証拠の与えられていないカテゴリの確率値が、変数Vを周辺化することで計算され、最大となる確率値が計算されたカテゴリに、入力された画像が分類されるようになる。

【0137】

なお、本技術は、上述した例の他にも、ユーザの選択行為を支援する意思決定支援装置、遺伝子解析や心理学解析等の社会科学分析を行う分析装置、センサからの入力情報の解析を行う解析装置、ユーザの嗜好等に応じた推薦を行う推薦装置、その他のエージェントシステム等に適用することができる。

【0138】

なお、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

【0139】

例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

【0140】

また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0141】

さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

【符号の説明】

【0143】

11 情報処理装置, 51 入力部, 52 制御部, 53 記憶部, 54 出力部, 71 経路探索部, 72 実行判定部, 73 検定実行部, 74 V字構

10

20

30

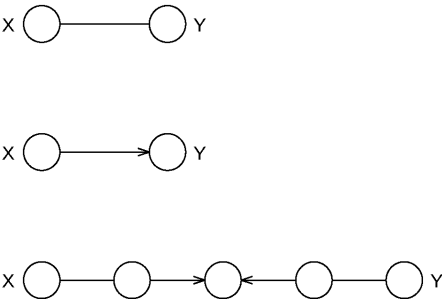
40

50

造探索部， 7 5 矛盾検出部， 8 1 変数対記憶部， 8 2 V字構造記憶部， 8
3 分離変数組記憶部

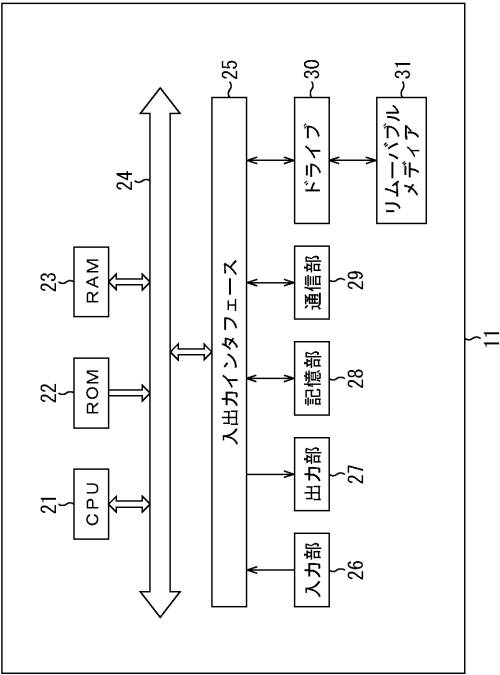
【図 1】

図1



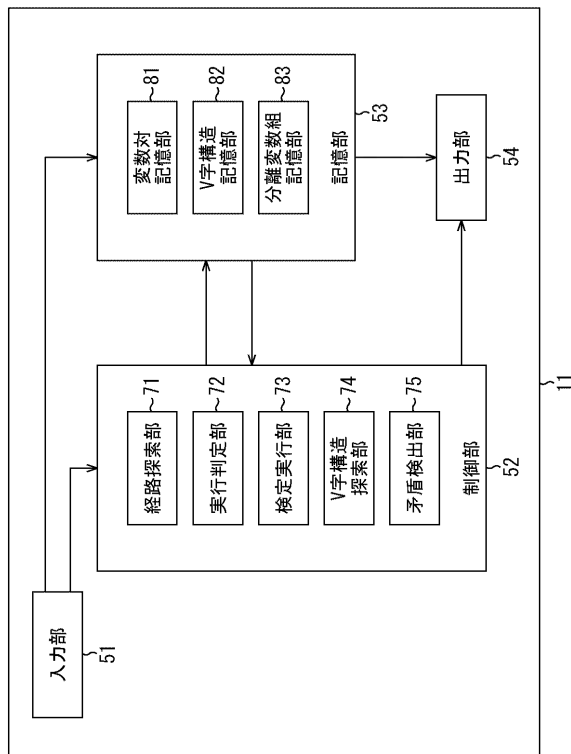
【図 2】

図2



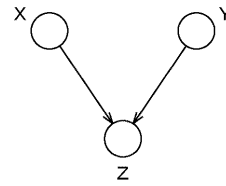
【図3】

図3



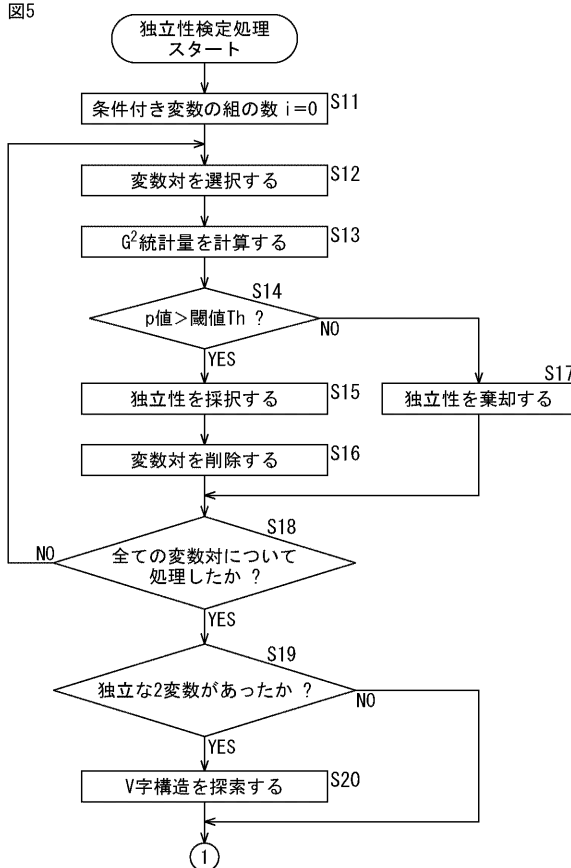
【図4】

図4



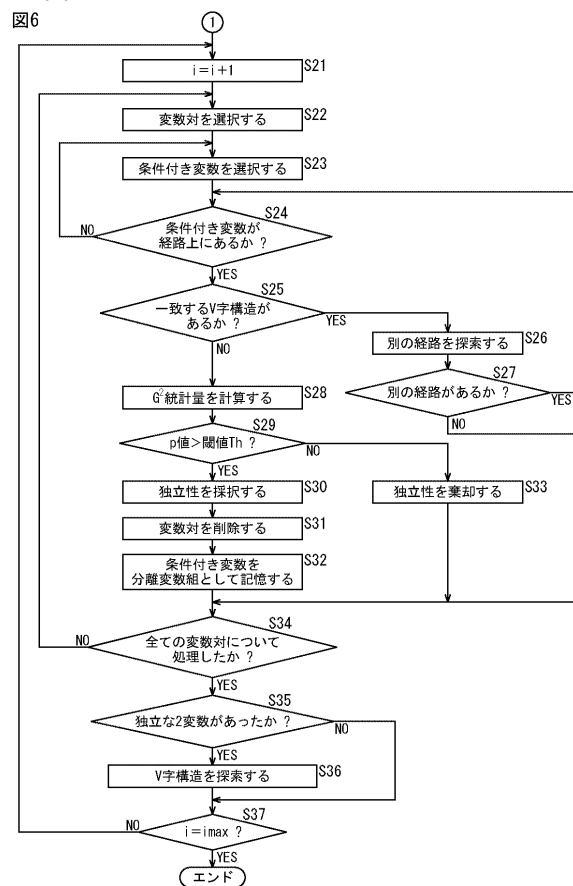
【図5】

図5



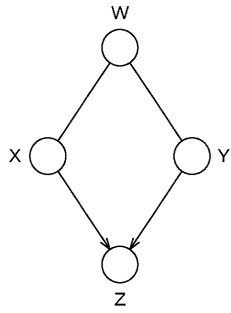
【図6】

図6



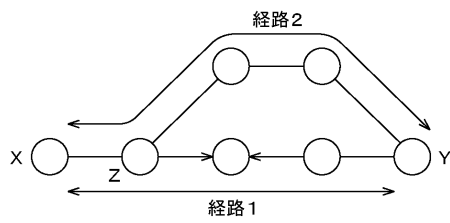
【図 7】

図7



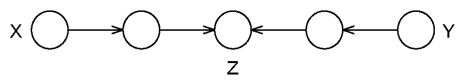
【図 8】

図8



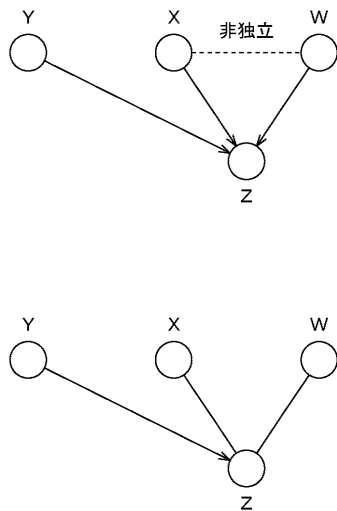
【図 9】

図9



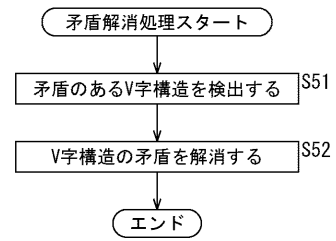
【図 12】

図12



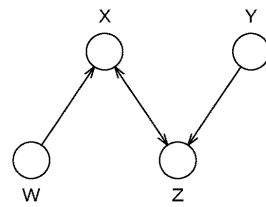
【図 10】

図10



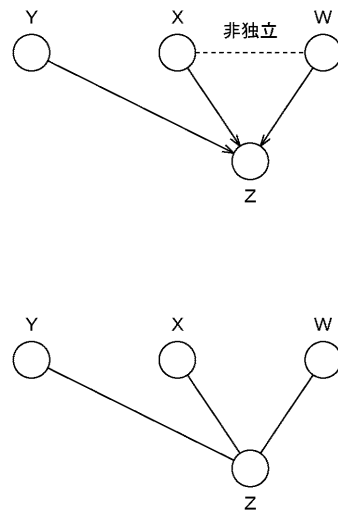
【図 11】

図11



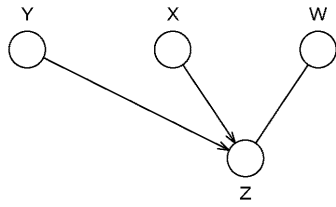
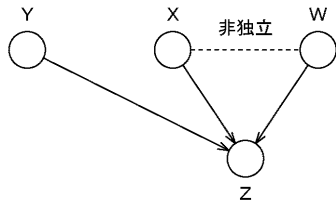
【図 13】

図13



【図 14】

図14



フロントページの続き

- (56)参考文献 特開 2011-048435 (JP, A)
特開 2007-066260 (JP, A)
米国特許出願公開第 2006/0047616 (US, A1)
磯崎 隆司, ベイジアンネットワークにおける因果発見, 人工知能学会誌, 日本, (社)人工知能学会, 2010年11月 1日, 第25巻 第6号, p. 811-818
森下 民平, Fast Constraint-based Learning of Bayesian Network using Frequent Itemset Mining, 第75回 人工知能基本問題研究会資料 (SIG-FPAI-A902), 日本, 社団法人人工知能学会, 2009年11月 6日, p. 5-11

(58)調査した分野(Int.Cl., DB名)

G06F 17/30
G06N 5/04