(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年7月30日(30.07.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/153150 A1

(51) 国際特許分類:
G06F 16/00 (2019.01)

(21) 国際出願番号： PCT/JP2020/000596

(22) 国際出願日： 2020年1月10日(10.01.2020)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2019-009760 2019年1月23日(23.01.2019) JP

(71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目５番１号 Tokyo (JP).

(72) 発明者: 橋本 悠香 (**HASHIMOTO, Yuka**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ知的財産センタ内 Tokyo (JP). 松尾 洋一 (**MATSUO, Yoichi**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ知的財産センタ内 Tokyo (JP).

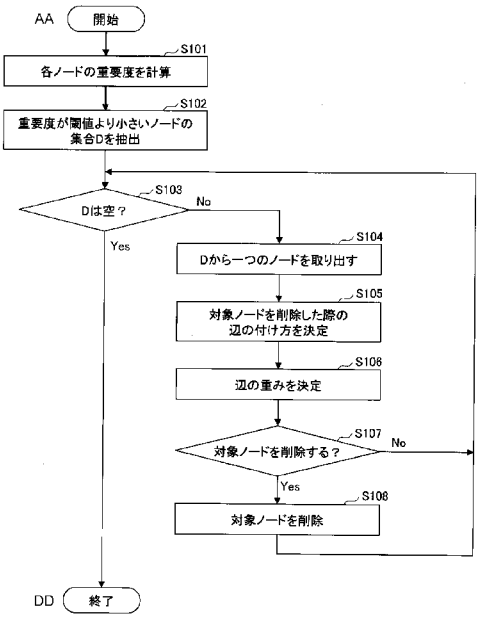
(74) 代理人: 伊東 忠重, 外 (**ITOH, Tadashige** et al.); 〒1000005 東京都千代田区丸の内二丁目１番１号 丸の内 ＭＹ ＰＬＡＺＡ （明治安田生命ビル） １６階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) **Title:** GRAPH SUMMARIZING DEVICE, GRAPH SUMMARIZING METHOD, AND PROGRAM

(54) 発明の名称： グラフ要約装置、グラフ要約方法及びプログラム

[図3]



S101 Calculate importance degree of each node
S102 Extract set D of nodes the importance degree of which is smaller than threshold
S103 Is D empty?
S104 Retrieve one node from D
S105 Determine how to add side when target node is deleted
S106 Determine weight of side
S107 Target node to be deleted?
S108 Delete target node
AA Start
DD End

(57) **Abstract:** A graph summarizing device comprising: a calculation unit which, when a graph, with each node having a cause degree that indicates the degree of cause for the state of the graph and each side having a weight that indicates the strength of causal relationship between the nodes, changes, calculates, for each node of the graph after change, the degree of importance based on the cause degree of nodes of the graph before change; a selection unit for selecting a first node the importance degree of which is smaller than or equal to a threshold as a candidate to delete; and a deletion unit for deleting

WO 2020/153150 A1

［続葉有］

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

---

the first node, whereby the graph summarizing device realizes graph summarization with which it is possible to suppress a decrease in the accuracy of cause estimation by a causal graph.

(57) 要約:グラフ要約装置は、各ノードがグラフの状態に対する要因の度合いを示す要因度を有し、各辺がノード間の因果関係の強さを示す重みを有するグラフが変化すると、変化後の前記グラフの各ノードについて、変化前の前記グラフのノードの要因度に基づく重要度を計算する計算部と、前記重要度が閾値以下である第1のノードを削除の候補として選択する選択部と、前記第1のノードを削除する削除部と、を有することで、因果グラフによる要因推定の精度の低下を抑制可能なグラフ要約を実現する。

明　細　書

発明の名称 ： グラフ要約装置、グラフ要約方法及びプログラム

技術分野

[0001] 本発明は、グラフ要約装置、グラフ要約方法及びプログラムに関する。

背景技術

[0002] システム運用において、オペレータの運用負荷低減のため、故障対応の自動化技術が検討されている。特に、障害発生時に障害の要因を特定することは人手では膨大な時間を要する場合が多く、障害要因推定技術に関する研究は重要である。これらの技術はＢａｙｅｓｉａｎ　Ｎｅｔｗｏｒｋ（非特許文献１）や決定木（非特許文献２）などの因果グラフを用いて行われるケースが多い。

[0003] 将来的には、仮想化技術の浸透により、システムの大規模化・複雑化が見込まれる。このため、障害発生時のアラートや異常なイベントの数が増加し、それらの因果関係も複雑化すると見込まれる。これにより、上記のような要因推定技術において、大規模かつ複雑なグラフが生じると見込まれる。

[0004] グラフに関する一般的な技術として、グラフの重要な部分を抜き出して単純化する、グラフ要約の技術が提案されている。これを因果グラフに適用できれば、グラフ要約を用いて要約後のグラフで要因推定を行い、おおまかな要因を推定した上で、要因となったイベントの付近のみについて要約前の情報を戻し、もう一度要因推定を行うことで計算時間の削減が期待できる。要因推定の対象となるイベント数が削減されることで計算時間が削減され、さらに、２段階の推定により、精度も確保される。

[0005] グラフ要約に関しては、個人又はコミュニティ同士の関係性や論文の引用関係の解析を対象とした研究が行なわれてきたが（非特許文献３）、要因推定を目的として因果グラフに対して適用するための技術は少ない。非特許文献４では、構造的な情報からＢａｙｅｓｉａｎ　Ｎｅｔｗｏｒｋを単純化することで、推定を効率化する技術が提案されている。

先行技術文献

非特許文献

[0006] 非特許文献1：E. R. J. Hruschka, M. do Carmo Nicoletti, V. A. de Oliveira, and G. M. Bressan, "Markov-blanket based strategy for translating a bayesian classifier into a reduced set of classification rules," in Proceedings of the 7th International Conference on Hybrid Intelligent Systems (HIS '07), pp. 192-197, 2007.

非特許文献2：A. X. Zheng, J. Lloyd, and E. Brewer, "Failure diagnosis using decision trees," in proceedings of the First International Conference on Autonomic Computing (ICAC '04), pp. 36-43, 2004.

非特許文献3：Y. Liu, T. Safavi, A. Dighe and D. Koutra, "Graph Summarization Methods and Applications: A Survey, ACM Computing Surveys," 51(3), pp. 1-34, 2018.

非特許文献4：M. Shiba, A. Takahashi, S. Aoki, H. Tsuji and S. Inoue, "Numerical experimentation on structure simplification in Bayesian network," 2009 IEEE International Conference on Systems, Man and Cybernetics, pp. 4698-4703, 2009.

発明の概要

発明が解決しようとする課題

[0007] 因果グラフに対してグラフ要約を適用することを考慮した方法は少ないながら存在するが、いずれの方法も、末端にあるノードを削除する、隣接ノードを共有している複数のノードをグルーピングするなど、グラフの構造的な情報しか用いていない。さらに、指針のみが示されており、削除するノードやグルーピングするノードの決定方法が明確に示されているわけではない。

[0008] 一方、システム運用における各ノードの要因へのなりやすさは、グラフの構造的な情報に加えて、過去の障害においてそのノードが要因であったか否かの情報も考慮する必要があるため、従来技術では因果グラフの要因推定の

精度が落ちてしまうという課題がある。　また、個人又はコミュニティ同士の関係性や論文の引用関係を対象とした技術は、それらの解析に特化した方法であるため、ノードの次数など各ノードの影響力を重視した要約を行う。

[0009] 一方、因果グラフによる要因推定においては、各ノードの要因へのなりやすさが重要となるため、上記のような既存技術でも因果グラフの要因推定の精度が落ちてしまうという課題がある。

[0010] このように、システム運用における因果グラフによる要因推定に適したグラフ要約は、既存技術のみでは達成することができない。

[0011] 本発明は、上記の点に鑑みてなされたものであって、因果グラフによる要因推定の精度の低下を抑制可能なグラフ要約を実現することを目的とする。

課題を解決するための手段

[0012] そこで上記課題を解決するため、グラフ要約装置は、各ノードがグラフの状態に対する要因の度合いを示す要因度を有し、各辺がノード間の因果関係の強さを示す重みを有するグラフが変化すると、変化後の前記グラフの各ノードについて、変化前の前記グラフのノードの要因度に基づく重要度を計算する計算部と、前記重要度が閾値以下である第１のノードを削除の候補として選択する選択部と、前記第１のノードを削除する削除部と、を有する。

発明の効果

[0013] 因果グラフによる要因推定の精度の低下を抑制可能なグラフ要約を実現することができる。

図面の簡単な説明

[0014] [図1]本発明の実施の形態における障害要因推定装置１０のハードウェア構成例を示す図である。

[図2]本発明の実施の形態における障害要因推定装置１０の機能構成例を示す図である。

[図3]障害要因推定装置１０が実行する処理手順の一例を説明するためのフローチャートである。

[図4]重みの決定の説明に用いるグラフ構造の第１の例を示す図である。

[図5]重みの決定の説明に用いるグラフ構造の第２の例を示す図である。

[図6]本実施の形態の評価における要約前のグラフを示す図である。

[図7]本実施の形態の評価における要約後のグラフを示す図である。

[図8]要約後のグラフの任意のノード間の要因度の差のヒートマップを示す図である。

発明を実施するための形態

[0015] 以下、図面に基づいて本発明の実施の形態を説明する。計算時間の増大と精度の低下という課題解決のため、本実施の形態では因果グラフの要約において重要なポイントを考慮したグラフ要約方法と、当該グラフ要約方法を実行する障害要因推定装置１０とが開示される。障害要因推定装置１０は、監視対象のＩＣＴ（Information and Communication Technology）システム（以下、単に「システム」という。）の障害の発生に応じて以下のフローを実行する。

[0016] ステップ１：現在（現在から過去の一定期間）においてシステムから出力されたログデータからイベントを抽出。ログデータから異常なイベントを抽出する場合、ログのテンプレート化を用いて異常なイベントが抽出されてもよい。また、システムが、明示的に異常なイベント（アラート）を出力する場合、ログデータからイベントを抽出するのではなく、各種類のアラートが、異常なイベントとして抽出されてもよい。なお、ログのテンプレート化については、例えば、「T. Kimura et. al., ”Spatio-temporal Factorization of Log Data for Understanding Network Events,” IEEE INFOCOM 2014, pp. 610-618, 2014.」等が参考とされてもよい。

[0017] ステップ２：過去の類似障害におけるイベント情報と、ステップ１において抽出されたイベントのイベント情報とを組み合わせて、イベントの因果関係を表す因果グラフを作成。イベントの因果関係を示す因果グラフの作成については、例えば、「P. Chen, Y. Qi, P. Zheng and D. Hou, ”CauseInfer: Automatic and Distributed Performance Diagnosis with Hierarchical Causality Graph in Large Distributed Systems,” IEEE INFOCOM 2014, pp. 18

87-1895, 2014.」、あるいは、「B. Zong, Y. Wu, J. Song, A. K. Singh, H. Cam, J. Han, and X. Yan, "Towards scalable critical alert mining," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14), pp.1057-1066, 2014.」等が参考とされてもよい。

[0018] ステップ３：ステップ２で作成したグラフを要約。

[0019] ステップ４：要約後のグラフで要因推定。要因推定については、例えば、「B. Zong, Y. Wu, J. Song, A. K. Singh, H. Cam, J. Han, and X. Yan, "Towards scalable critical alert mining," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14), pp.1057-1066, 2014.」等が参考とされてもよい。

[0020] ステップ５：ステップ４で要因となったイベントの付近のみについて要約前の情報を戻し、もう一度要因推定。これにより、障害要因が推定される。

[0021] 本実施の形態では、上記のステップ３のグラフ要約技術に焦点を当て、新技術を提案する。要因推定に用いる因果グラフの要約において重要なポイントとなるのは次の３つである。

（１）要因となる可能性が高いノードは要約後も残す。

（２）要約の前後で要因推定の結果に変化が少ない。

（３）要約後のグラフはサイクル構造を持たない。

[0022] 本実施の形態の障害要因推定装置１０は、上記（１）～（３）のポイントを考慮することで、因果グラフによる要因推定において適用可能なグラフ要約を行う。本実施の形態は以下のフローで実行される。

[0023] ステップ１：ポイント（１）を考慮し重要度の低いノードを検出

ステップ２：ポイント（２）を考慮し、該当ノード削除後のグラフの辺及び重みを決定

ステップ３：ポイント（３）を考慮し、検出したノードを削除するか否かを判断

［問題設定］

因果グラフGは、G＝（V（G），E（G），f_G）とする。但し、V（G）＝｛v_1，…，$v_{N(G)}$｝は、グラフGのノードの集合で、各ノードは障害時に発生したアラート、又は障害時にログデータから抽出したイベントを表し、それぞれが要因度を有する。要因度とは、因果グラフGの現在の状態（構造）に対する要因である度合いを示す指標である。なお、本実施の形態では、障害時に発生したアラート及びログデータから抽出したイベントのいずれをも、イベントと呼ぶ。N（G）は、グラフGのノード数、E（G）＝｛$e_{j,k}$｜j，k＝1，…，N（G）｝は、グラフGの辺の集合で、$e_{j,k}$は、ノードv_jからノードv_kへ向かう辺を表す。f_G：E（G）→R_+は、Gの辺の重みを表す関数とする。ただし、R_+は0以上の実数全体の集合を表す。辺の重みは、当該辺によって接続されるノード間の因果関係の強さを示す。すなわち、辺の重みは、当該辺の先のノードに係るイベントの要因が、当該辺の元のノードに係るイベントである可能性の高さを示す。また、Gの隣接行列A（G）を、A（G）＝［f_G（$e_{j,k}$）］$_{j,k}$で定める。

[0024] システム運用において、最初の障害が起こると、因果グラフG_1が生成される。次に類似した障害が起こった場合、1度目の障害の情報に2度目の障害の情報が加わり、V（G_1）⊆V（G_2）を満たす因果グラフG_2が生成される（グラフG_1の状態（構造）がグラフG_2に変化する）。障害が起こる度にこれが行われ、類似障害に対するグラフの列｛G_i｝$_{i=1}^{\infty}$が生成される。また、i度目の障害時に発生したイベントの集合をR_i⊆V（G_i）とし、集合列｛R_i｝$_{i=1}^{\infty}$が生成されるとする。iが増加するにしたがって、イベント間の因果関係は定まり、真の因果関係を表す因果グラフへ収束していくとする。また、iが増加するにしたがって各発生イベントがノイズかどうかの判断が可能となり、真の発生イベントへ収束するとする。つまり、次を満たすG＝（V（G），E（G），f_G）とRが存在すると仮定する。

[0025]

[数1]

$$\lim_{i\to\infty} V(G_i) = V(G), \qquad \lim_{i\to\infty} E(G_i) = E(G), \qquad \lim_{i\to\infty} f_{G_i} = f_G \text{ (各点)}$$

$$\lim_{i\to\infty} R_i = R$$

但し、集合列$\{S_i\}_{i=1}^{\infty}$に対して、上極限$\bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty}S_i$と、下極限$\bigcup_{k=1}^{\infty}\bigcap_{i=k}^{\infty}S_i$が一致するとき、$\bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty}S_i=\bigcup_{k=1}^{\infty}\bigcap_{i=k}^{\infty}S_i=\mathrm{lim}_{i\to\infty}S_i$と表すこととする。

[0026] 極限グラフGにおいて、確率空間（Ω，Ｆ，ｐ）が定まるとする。但し、Ωは、Ｖ（Ｇ）の全ての部分集合からなる集合族とする。また、各$v_j\in V(G)$に対して確率変数$X_j:\Omega\to\{0,1\}$を、$v_j\in\omega$のとき$X_j(\omega)=1$、$v_j\notin\omega$でないとき$X_j(\omega)=0$で定める。$f_G(e_{j,k})=p(\{X_k=1\}\mid\{X_j=1\})$とする。各ｉに対して$f_{G_i}$は、$f_G$を近似する関数で、既存手法により定める。

[0027] このとき、イベント$v_j\in V(G_i)$が起こった際に、イベント$v_l\in R_i$が起こる確率の和$\Sigma_{v_l\in R_i}p(\{X_l=1\}\mid\{X_j=1\})$の近似$P_i(v_j;R_i)$を$v_j$の要因度とし、この値が高いイベントが、ｉ度目の障害時の主要因であると推定することとする。但し、P_iは次のように定める。

[0028] [数2]

$$P_i(v;\{v\}) = 1,$$

$$P_i(v_k;R_i) = \sum_{v\in R_i}\left(1-\prod_{e_{k,l}\in E(G_i)}\left(1-P_i(v_l;\{v\})f_{G_i}(e_{k,l})\right)\right)$$

この数式に関しては、例えば、「B. Zong, Y. Wu, J. Song, A. K. Singh,

H. Cam, J. Han, and X. Yan, "Towards scalable critical alert mining, " in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14), pp.1057-1066, 2014.」等が参考とされてもよい。

[0029] このとき、$\lim_{i\to\infty} P_i(\cdot;R_i) = P(\cdot;R)$（各点）が成り立つ。但し、

[0030] [数3]

$$P(v;\{v\}) = 1,$$

$$P(v_k;R) = \sum_{v\in R}\left(1 - \prod_{e_{k,l}\in E(G)}\left(1 - P(v_l;\{v\})f_G(e_{k,l})\right)\right)$$

である。

[0031] 類似障害が十分な回数起これば、$G_*\fallingdotseq G$、$R_*\fallingdotseq R$を満たすG_*、R_*が得られ、そこから$P_*(\cdot;R_*)$を計算することでその故障に対する主要因が確定する。しかし、G_*、R_*が得られるまでは、その時点で得られている情報を用いて要因推定を行い、出来るだけ高速、かつ、正確に主要因を推定する必要がある。そこで、G_1，…，G_{i-1}の情報を用いてG_iを要約し、$V(G'_i)\subset V(G_i)$を満たすG'_iを作成した上でG'_iに対して要因推定を行うことで、上記を達成することを考える。

[0032] ［障害要因推定装置１０］

上記考えに基づく障害要因推定装置１０について説明する。図１は、本発明の実施の形態における障害要因推定装置１０のハードウェア構成例を示す図である。図１の障害要因推定装置１０は、それぞれバスＢで相互に接続されているドライブ装置１００、補助記憶装置１０２、メモリ装置１０３、ＣＰＵ１０４、及びインタフェース装置１０５等を有するコンピュータである。

[0033] 障害要因推定装置１０での処理を実現するプログラムは、ＣＤ－ＲＯＭ等の記録媒体１０１によって提供される。プログラムを記憶した記録媒体１０１がドライブ装置１００にセットされると、プログラムが記録媒体１０１からドライブ装置１００を介して補助記憶装置１０２にインストールされる。但し、プログラムのインストールは必ずしも記録媒体１０１より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置１０２は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

[0034] メモリ装置１０３は、プログラムの起動指示があった場合に、補助記憶装置１０２からプログラムを読み出して格納する。ＣＰＵ１０４は、メモリ装置１０３に格納されたプログラムに従って障害要因推定装置１０に係る機能を実行する。インタフェース装置１０５は、ネットワークに接続するためのインタフェースとして用いられる。

[0035] 図２は、本発明の実施の形態における障害要因推定装置１０の機能構成例を示す図である。図２において、障害要因推定装置１０は、検出部１１、決定部１２及び削除部１３を有する。これら各部は、障害要因推定装置１０にインストールされた１以上のプログラムが、ＣＰＵ１０４に実行させる処理により実現される。

[0036] 図３は、障害要因推定装置１０が実行する処理手順の一例を説明するためのフローチャートである。

[0037] ステップＳ１０１において、検出部１１は、ｉ度目の障害の発生に応じた変化後の因果グラフG_iにおける各ノードの重要度を計算する。

[0038] G_iにおいて要因推定を行う際、$b_i(v)=P_{i-1}(v;R_{i-1})$（$v\in V(G_{i-1})\subseteq V(G_i)$のとき）、$b_i(v)=0$（$v\in V(G_i)\setminus V(G_{i-1})$のとき）と定める。なお、$G_{i-1}$は、ｉ度目の障害の発生前（すなわち、変化前の）因果グラフである。

[0039] $b_i(v)$が大きい、つまり、G_{i-1}において要因度が大きかったイベントｖは、G_iでも要因度が高くなると見込まれるが、新たに加わったイベントｕ

∈Ｖ（Ｇ$_{i}$）＼Ｖ（Ｇ$_{i-1}$）については、当該イベントｕがＧ$_{i-1}$において要因度が大きかったイベントｖの要因となり得る場合、要因度が高くなる可能性がある。但し、Ｇ$_{i}$における各ノードの要因度の計算の負荷は高い。そこで、検出部１１は、要因度の近似値であって（要因度の大小関係を近似できる値であって）、要因度よりも低い計算負荷によって計算可能な重要度を、Ｇ$_{i-1}$における要因度を用いてＧ$_{i}$の各ノードについて計算する。

[0040] ｖ∈Ｖ（Ｇ$_{i}$）の重要度をｃ$_{i}$（ｖ）とおき、ｃ$_{i}$＝［ｃ$_{i}$（ｖ$_{1}$），…，ｃ$_{i}$（ｖ$_{N(Gi)}$）］^{T}、と定める。さらに、ｂ$_{i}$＝［ｂ$_{i}$（ｖ$_{1}$），…，ｂ$_{i}$（ｖ$_{N(Gi)}$）］^{T}と定める。重要度は、時刻ｔ＝０においてｂ$_{i}$であり、時刻ｔ＝１においてｃ$_{i}$になるとする。ｂ$_{i}$とｃ$_{i}$を時刻ｔについて滑らかにつなぐ関数ｄ$_{i}$：［０，Ｔ］→Ｒ$_{+}^{N(Gi)}$を考える。ｖ$_{j}$を、Ｇ$_{i}$において新たに加わったイベントとする。微小時ｔ＝０からｔ＝ｈにおけるｄ$_{i}$の第ｊ成分の変化量を考える。ｖ$_{j}$に隣接しているノードの集合を

[0041] [数4]

$$\{u_1, \dots, u_{d_{G_i}(v_j)}\}$$

とする。ｔ＝ｈでのｄ$_{i}$の第ｊ成分は、ｕ$_{k}$（ｋ＝１，…，ｄ$_{Gi}$（ｖ$_{j}$））の各要因度にｐ（｛Ｘ$_{k}$＝１｝｜｛Ｘ$_{j}$＝１｝）を掛け合わせ、ｋについて足し合わせたものと考える。これにより、Ｇ$_{i}$における要因度が相対的に高くなる可能性があるイベントの重要度が相対的に低くならないようにすることができる。

[0042] 具体的には、検出部１１は、次のようにしてＧ$_{i}$における重要度ｃ$_{i}$：Ｖ（Ｇ$_{i}$）→Ｒ$_{+}$を計算する。

[0043]

[数5]

$$\frac{d}{dt}\boldsymbol{d_i} = A_i\boldsymbol{d_i}, \qquad \boldsymbol{d_i}(0) = \boldsymbol{b_i}$$

$$\boldsymbol{c_i} \coloneqq \boldsymbol{d_i}(1) = e^{A_i}\boldsymbol{b_i} \qquad (1)$$

但し、A_iは、G_iの隣接行列A（G_i）を表す。

[0044] 後述されるように、検出部１１は、（１）式により計算した重要度c_i（ｖ）が相対的に小さくなるｖを、要因推定においてあまり重要ではないとみなし、削除の候補として選択する。以下が成立する。

[0045] [数6]

$$\begin{aligned}
c_i(v_j) = d_{i,j}(1) &= d_{i,j}(0) + \int_0^1 (A_i\boldsymbol{d_i})_j\, dt \\
&\geq \int_0^1 \sum_{e_{jk}\in E(G_i)} d_{i,k}(t) f_{G_i}(e_{j,k}) dt \\
&= \sum_{e_{jk}\in E(G_i)} \int_0^1 d_{i,k}(t) dt\, f_{G_i}(e_{j,k}) \\
&\geq \sum_{e_{jk}\in E(G_i)} d_{i,k}(0)\, f_{G_i}(e_{j,k}) \\
&\geq \sum_{e_{jk}\in E(G_i)} \sum_{v\in R_{i-1}} P_{i-1}(v_k;\{v\})\, f_{G_i}(e_{j,k}) \\
&\geq \sum_{v\in R_{i-1}} \left(1 - \prod_{e_{jk}\in E(G_i)} (1 - P_{i-1}(v_k;\{v\})\, f_{G_i}(e_{j,k}))\right) \\
&\xrightarrow[i\to\infty]{} \sum_{v\in R} \left(1 - \prod_{e_{jk}\in E(G)} (1 - P(v_k;\{v\})\, f(e_{j,k}))\right) = P(v_j;R)
\end{aligned}$$

但し、４行目の不等式は、A_iの成分は全て正であるため、d_iの各成分はｔ

に関して単調増加することを用いた。また、$(A_i d_i)_j$は、$A_i d_i$の第ｊ成分を表す。よって、重要度$c_i(v)$は、$i \to \infty$において、真の要因度Ｐ（ｖ；Ｒ）以上の値になる。

[0046] また、d_iの各成分は単調増加するため、任意の$v \in V(G_i)$に対して、$b_i(v) \leqq c_i(v)$が成り立つから、G_{i-1}において要因度が相対的に高かったノードが削除されることはない。これにより、ポイント（１）を考慮した検出が可能となる。

[0047] 続いて、検出部１１は、$c_i(v)$が一定の閾値c_{max}以下であるノードの集合を抽出し、抽出した集合をＤに代入する（Ｓ１０２）。すなわち、以下の演算が行われる。

$D = \{v \in V(G_i) \mid c_i(v) \leqq c_{max}\}$

続いて、決定部１２は、該当ノード（Ｄに含まれるノード）の有無（Ｄが空であるか否か）を判定する（Ｓ１０３）。該当ノードが無い場合（Ｄが空である場合）（Ｓ１０３でＹｅｓ）、図３の処理は終了する。該当ノードが有る場合（Ｓ１０３でＮｏ）、決定部１２は、Ｄから一つのノードを取り出す（Ｓ１０４）。取り出されたノードは、Ｄから削除される。以下、取り出されたノードを、「ノードv_j」という。なお、ノードの取り出し順は、特定の順序に限定されない。例えば、重要度$c_i(v)$が最も低いノードが取り出されてもよいし、インデックスが最も小さいノードが取り出されてもよいし、ランダムにノードが取り出されてもよい。インデックスとは、ノードに付与される識別子であり、例えば、ノードが生成された順番を示す数値である。

[0048] 続いて、決定部１２は、ノードv_jを削除した際（削除した後）の辺の付け方を決定する（Ｓ１０５）。具体的には、決定部１２は、ノードv_jに隣接するノードのうち、v_jからの辺の重みが最大になるノードv_kを特定し、v_jとv_kを１つのノードとみなすことにより、当該１つのノードにv_jの辺及びv_kの辺を付ける。

[0049] 重みが最大になるノードを用いる理由は、以下で説明する方法で重みを決

定した場合に、ノード削除前と削除後の要因度の計算における重みの変化の影響を最小にするためである。

[0050] 続いて、決定部１２は、ステップＳ１０５において付け方が決定された各辺の重みを、次の３つの単純な構造において、ノード削除前後で各ノードの要因度が変化しないように決定する（Ｓ１０６）。以下では、ノードv_j削除前のグラフをG_B、削除後のグラフをG_Aと表す。

[0051] ［木構造］

共通のノードからv_jとv_kの両方に辺が入っておらず、かつ、共通のノードにv_jとv_k両方から辺がのびていない場合（すなわち、ノードv_jがノードv_kの親ノードである場合）、v_jからv_kへ向かう辺の重み$f_{GB}(e_{jk})$を、v_kから出ている辺の重み全てに掛け合わせることにより、f_{GA}を作成する。このようにして重みを決定することで、木に対してはノード削除前後で要因度が変化しないようにすることができる。

[0052] ［図４の構造］

G_Bにおいて共通のノードv_lからv_jとv_kの両方に辺が入っている場合、ステップＳ１０５ではe_{jk}、e_{lk}、e_{lj}の３辺が、G_Aにおいて一つの辺e_{lk}となる。重みは、$f_{GA}(e_{lk}) = f_{GB}(e_{lj}) \cdot f_{GB}(e_{jk}) + f_{GB}(e_{lk}) - f_{GB}(e_{lj}) \cdot f_{GB}(e_{jk}) \cdot f_{GB}(e_{lk})$により決定する。このように決定することで、図４のような構造に対してはノード削除前後で要因度が変化しないようにすることができる。なお、図４において、$a = f_{GB}(e_{lj})$、$b = f_{GB}(e_{lk})$、$c = f_{GB}(e_{jk})$である。

[0053] ［図５の構造］

G_Bにおいて共通のノードv_lに、v_jとv_kの両方から辺がのびている場合、ステップＳ１０５ではe_{jk}、e_{kl}、e_{jl}の３辺が、G_Aにおいて一つの辺e_{jl}となる。重みは、$f_{GA}(e_{kl}) = f_{GB}(e_{kl}) \cdot f_{GB}(e_{jk}) + f_{GB}(e_{jl}) - f_{GB}(e_{kl}) \cdot f_{GB}(e_{jk}) \cdot f_{GB}(e_{jl})$により決定する。このように決定することで、図５のような構造に対してはノード削除前後で要因度が変化しないようにすることができる。なお、図５において、図４に

おいて、$a = f_{GB}(e_{jk})$、$b = f_{GB}(e_{jl})$、$c = f_{GB}(e_{kl})$である。

[0054] 続いて、削除部１３は、ノードv_jを削除するか否かを判断する（Ｓ１０７）。具体的には、削除部１３は、削除候補のv_jの削除前のグラフをG_B、削除後のグラフをG_Aとした場合にG_Aの隣接行列の固有値を調べることで、G_Aにサイクル構造があるか否かを調べる。一般に、グラフＧの隣接行列Ａ（Ｇ）に対して、次が成り立つ。

Ａ（Ｇ）の固有値が全て０

⇒$A(G)^n = 0$

⇒Ｇに長さｎ以上のパスは無い。

⇒Ｇにサイクル構造が無い。

[0055] よって、削除部１３は、Ａ（G_A）の固有値が全て０ならば（Ｓ１０８でＹｅｓ）、v_jを削除し、そうでなければ（Ｓ１０８でＮｏ）、v_jを削除しない。なお、削除部１３は、v_jを削除した場合、ステップＳ１０５及びＳ１０６の実行結果に従って、G_iに対する辺の付与及び当該辺に対する重みを付与する。

[0056] ステップＳ１０４以降は、Ｄが空になるまで実行される。すなわち、各$v_j \in D$に対してステップＳ１０４以降が実行される。

[0057] 一般の行列に対する固有値の計算は計算コストが高いが、次の理由により、今回の場合は計算コストが削減される。一般に、グラフＧにサイクル構造が無ければ、Ａ（Ｇ）は上三角行列にできる。よって、G_iは上三角行列と仮定されてよい。Ａ（G_B）からＡ（G_A）への変化は、v_jに隣接するノードのうち、v_jからの辺の重みが最大になるノードv_kに対応する行と列の変化と、ノードv_jに対応する行と列の削除である。この変化の中で上三角行列の構造に変化を与えるのはノードv_kに対応する行と列の変化のみであるから、G_Aの上三角成分以外の成分の数は少ない。よって、G_Aの固有値を数値的に求めるために必要なＱＲ分解に必要な計算量は少なくなる。

[0058] ［評価結果］

ランダムに作成した小さなグラフに対して本実施の形態を適用することで、上記のポイント（１）～（３）のうち、「（１）要因となる可能性が高いノードは要約後も残す」、「（２）要約の前後で要因推定の結果に変化が少ない」について評価を行った。

[0059] ［設定］

以下のようにして、［問題設定］で述べた状況を模擬し、評価を行った。

１．ノード数が２５のランダムなグラフG_1を作成。

２．$R_1=V(G_1)$として、G_1の各ノード$v\in V(G_1)$に関して、要因度$P_1(v;R_1)$を計算。

３．G_1にノードを付け加えノード数を５０にし、ランダムに辺を付け加えてG_2を作成。

４．図３に示した処理手順で要約を行い、要約後のグラフG'_2を作成。

５．評価１：本実施の形態を用いて要約を行った際に、要約前後で各ノードの要因度の違いが小さいということを確認することを目的とし、各u，$v\in V(G'_2)$に対し、$P_2(v;\{u\})$と、G'_2において計算した要因度$P'_2(v;\{u\})$との差$P_2(v;\{u\})-P'_2(v;\{u\})$を計算。

６．R_2を、R_1の元のうちのランダムに選んだ５つを、$V(G_2)\setminus R_1$の元からランダムに選んだ５つによって交換したものをR_2として設定。

７．評価２：本実施の形態を用いて要約を行った際に、要約前後で主な要因として抽出される上位５イベントに差異が少ないということを確認することを目的とし、$P_2(v;R_2)$が大きくなる上位５つの$v\in V(G_2)$と、$P'_2(v;R_2)$が大きくなる５つの$v\in V(G'_2)$を求め、両者を比較する。

[0060] ［結果］

要約前のグラフG_2の５０のノード数に対し、要約後のグラフG'_2のノード数は４０となった。要約前のグラフG_2を図６に示し、要約後のグラフG'_2を図７に示す。

[0061] 評価１：要約後のグラフG'_2の任意のノードu，$v\in V(G'_2)$に対してノードuのインデックスを縦軸、ノードvのインデックスを横軸にとった時

の｜P_2（ｖ；｛ｕ｝）－P'_2（ｖ；｛ｕ｝）｜の値をヒートマップで表すと図８のようになった。ほとんどのｕ，ｖのペアでは、｜P_2（ｖ；｛ｕ｝）－P'_2（ｖ；｛ｕ｝）｜の値は０．１以下となった。これは、R_2＝｛ｖ｝とした際、グラフ要約の前後でイベントｕの要因度の差がほぼ０．１以内で収まっていることを表している。

[0062] 評価２：P_2（ｖ；R_2）が大きくなる上位５つのｖ∈Ｖ（G_2）のインデックスと、P'_2（ｖ；R_2）が大きくなる５つのｖ∈Ｖ（G'_2）のインデックスとは表１のようになった。

[0063] [表1]

	$P_2(v;R_2)$	$P'_2(v;R_2)$
1番目	0	10
2番目	10	0
3番目	4	4
4番目	2	9
5番目	15	2

多少の順序の入れ替わりはあるものの、上位５つのうち４つのインデックスは両方に現れており、要約前後で要因推定の結果はほぼ一致していることが分かる。

[0064] 上述したように、本実施の形態によれば、過去の情報の伝播を因果グラフによる要因推定に適する形で行い、その結果を基に不要なノードを検出することでグラフを要約する。すなわち、過去の情報を用いて各ノードの重要度を決定することで、重要でない頂点を検出する。この重要度は真の要因度以上の値に収束することが示せるから、要因となる可能性が高いノードは要約後も残すことができる。その結果、システム運用における因果グラフによる要因推定の精度の低下を抑制可能なグラフ要約を実現することができる。

[0065] また、単純な構造において、ノード削除前後で要因度が変化しないように

辺の重みを決定することで、要約の前後で要因推定の結果に変化が少なくなる。さらに、隣接行列の固有値を計算することで、要約後のグラフはサイクル構造を持たないことが達成できる。隣接行列の構造を考慮すると、この固有値計算は少ない計算量で達成できる。

[0066] すなわち、障害によって得られたイベントの因果関係を表す因果グラフを、できるだけ重要なイベントを残し、かつ、要因推定の結果に与える影響を出来るだけ小さくするように、かつ、グラフの構造を保って要約することで、ある程度の正確性を保ったまま要因推定を高速化できる。さらに、要約後のグラフで要因度が高かったノードの周辺のみ、削除した頂点を元に戻して再度要因推定を行えば、さらなる正確性を達成できる。これにより、障害発生時の対応を、正確性を保ったまま高速化できる。

[0067] なお、本実施の形態において、障害要因推定装置１０は、因果グラフ要約装置の一例である。検出部１１は、計算部及び選択部の一例である。

[0068] 以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

[0069]

１０	障害要因推定装置
１１	検出部
１２	決定部
１３	削除部
１００	ドライブ装置
１０１	記録媒体
１０２	補助記憶装置
１０３	メモリ装置
１０４	ＣＰＵ
１０５	インタフェース装置
Ｂ	バス

請求の範囲

[請求項1] 各ノードがグラフの状態に対する要因の度合いを示す要因度を有し、各辺がノード間の因果関係の強さを示す重みを有するグラフが変化すると、変化後の前記グラフの各ノードについて、変化前の前記グラフのノードの要因度に基づく重要度を計算する計算部と、

前記重要度が閾値以下である第１のノードを削除の候補として選択する選択部と、

前記第１のノードを削除する削除部と、

を有することを特徴とするグラフ要約装置。

[請求項2] 前記第１のノードごとに、当該第１のノードと、当該第１のノードからの辺の重みが最大である第２のノードとを１つのノードとして、当該第１のノードが削除された場合に前記グラフに付ける辺を決定し、当該第１のノードの削除前後における要因度が変化しないように当該辺の重みを決定する決定部、

を有することを特徴とする請求項１記載のグラフ要約装置。

[請求項3] 前記削除部は、前記第１のノードを削除した場合に前記グラフにサイクル構造が発生しない場合に当該第１のノードを削除する、

ことを特徴とする請求項１又は２記載のグラフ要約装置。

[請求項4] 各ノードがグラフの状態に対する要因の度合いを示す要因度を有し、各辺がノード間の因果関係の強さを示す重みを有するグラフが変化すると、変化後の前記グラフの各ノードについて、変化前の前記グラフのノードの要因度に基づく重要度を計算する計算手順と、

前記重要度が閾値以下である第１のノードを削除の候補として選択する選択手順と、

前記第１のノードを削除する削除手順と、

をコンピュータが実行することを特徴とするグラフ要約方法。

[請求項5] 前記第１のノードごとに、当該第１のノードと、当該第１のノードからの辺の重みが最大である第２のノードとを１つのノードとして、

当該第１のノードが削除された場合に前記グラフに付ける辺を決定し、当該第１のノードの削除前後における要因度が変化しないように当該辺の重みを決定する決定手順、

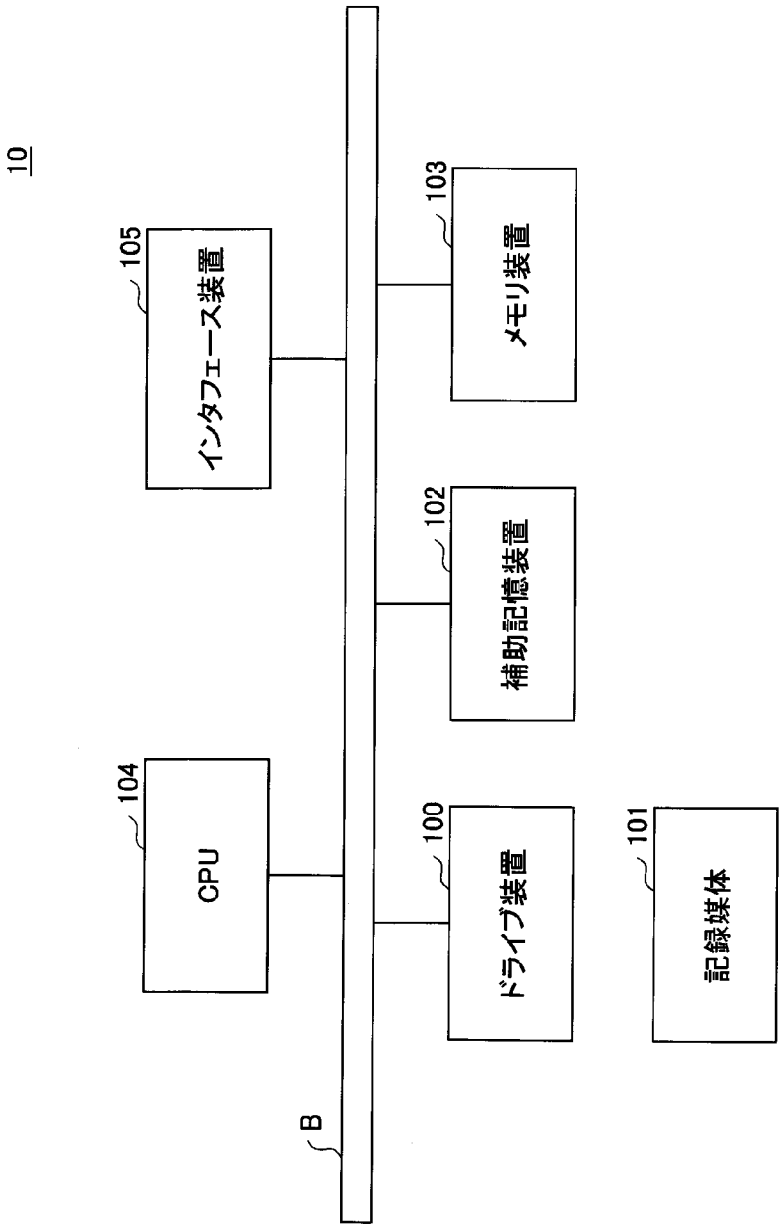
を前記コンピュータが実行することを特徴とする請求項４記載のグラフ要約方法。

[請求項6] 前記削除手順は、前記第１のノードを削除した場合に前記グラフにサイクル構造が発生しない場合に当該第１のノードを削除する、

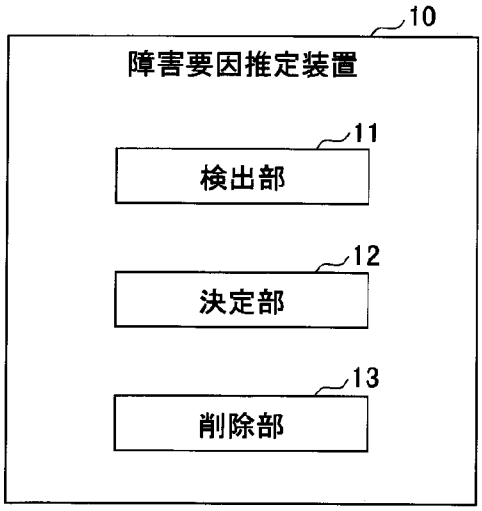
ことを特徴とする請求項４又は５記載のグラフ要約方法。

[請求項7] 請求項４乃至６いずれか一項記載のグラフ要約方法をコンピュータに実行させることを特徴とするプログラム。

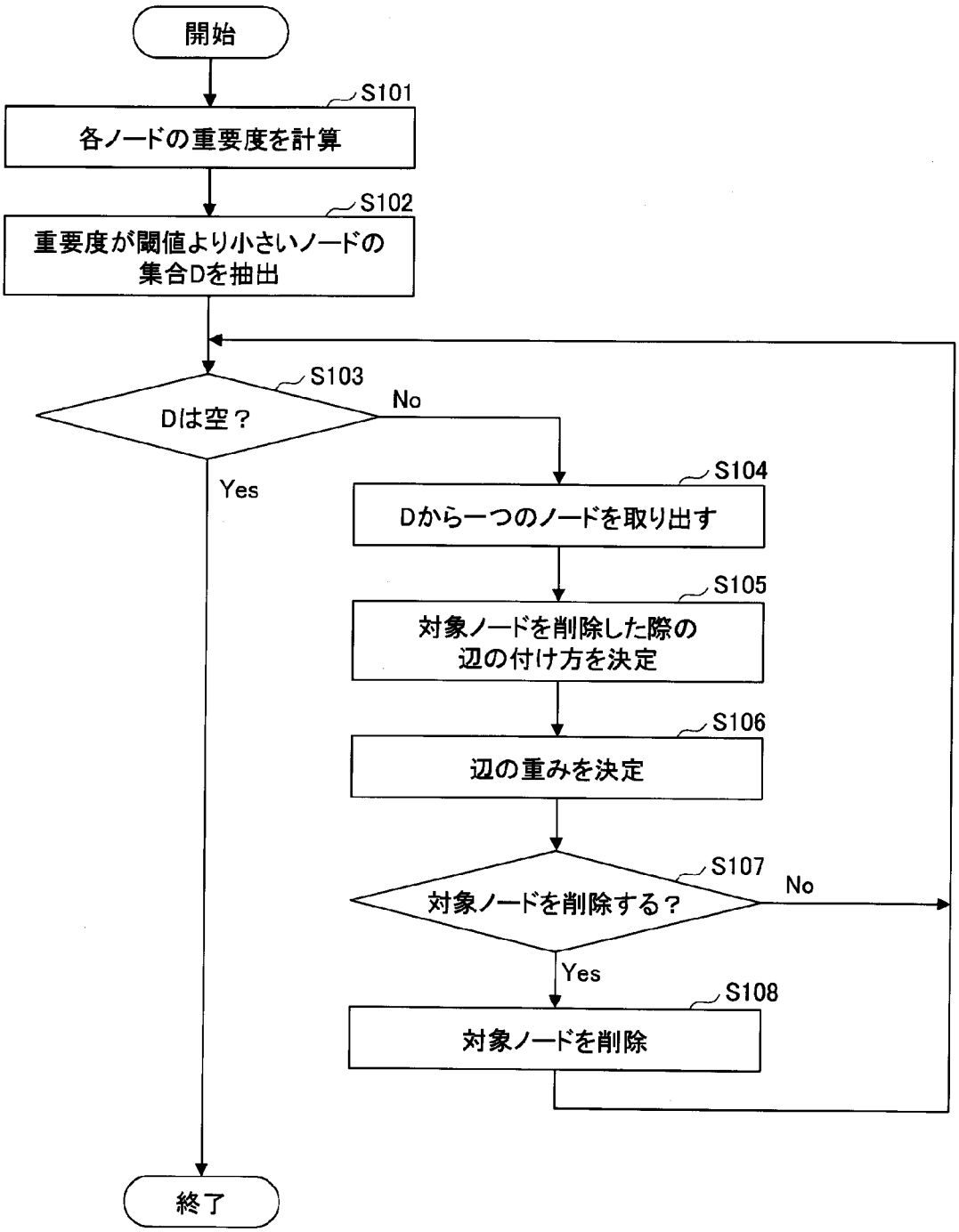
[図1]


10
105
インタフェース装置
103
メモリ装置
102
補助記憶装置
104
CPU
100
ドライブ装置
101
記録媒体
B

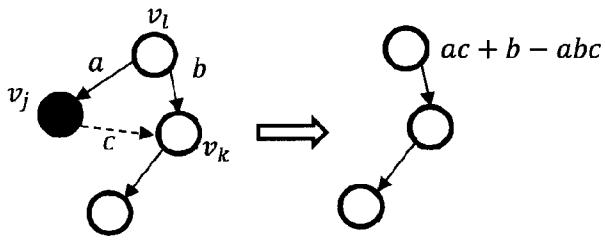
[図2]



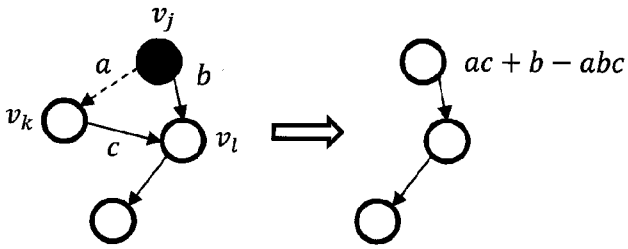
[図3]


開始
S101
各ノードの重要度を計算
S102
重要度が閾値より小さいノードの集合Dを抽出
S103
Dは空？
No
Yes
S104
Dから一つのノードを取り出す
S105
対象ノードを削除した際の辺の付け方を決定
S106
辺の重みを決定
S107
対象ノードを削除する？
No
Yes
S108
対象ノードを削除
終了

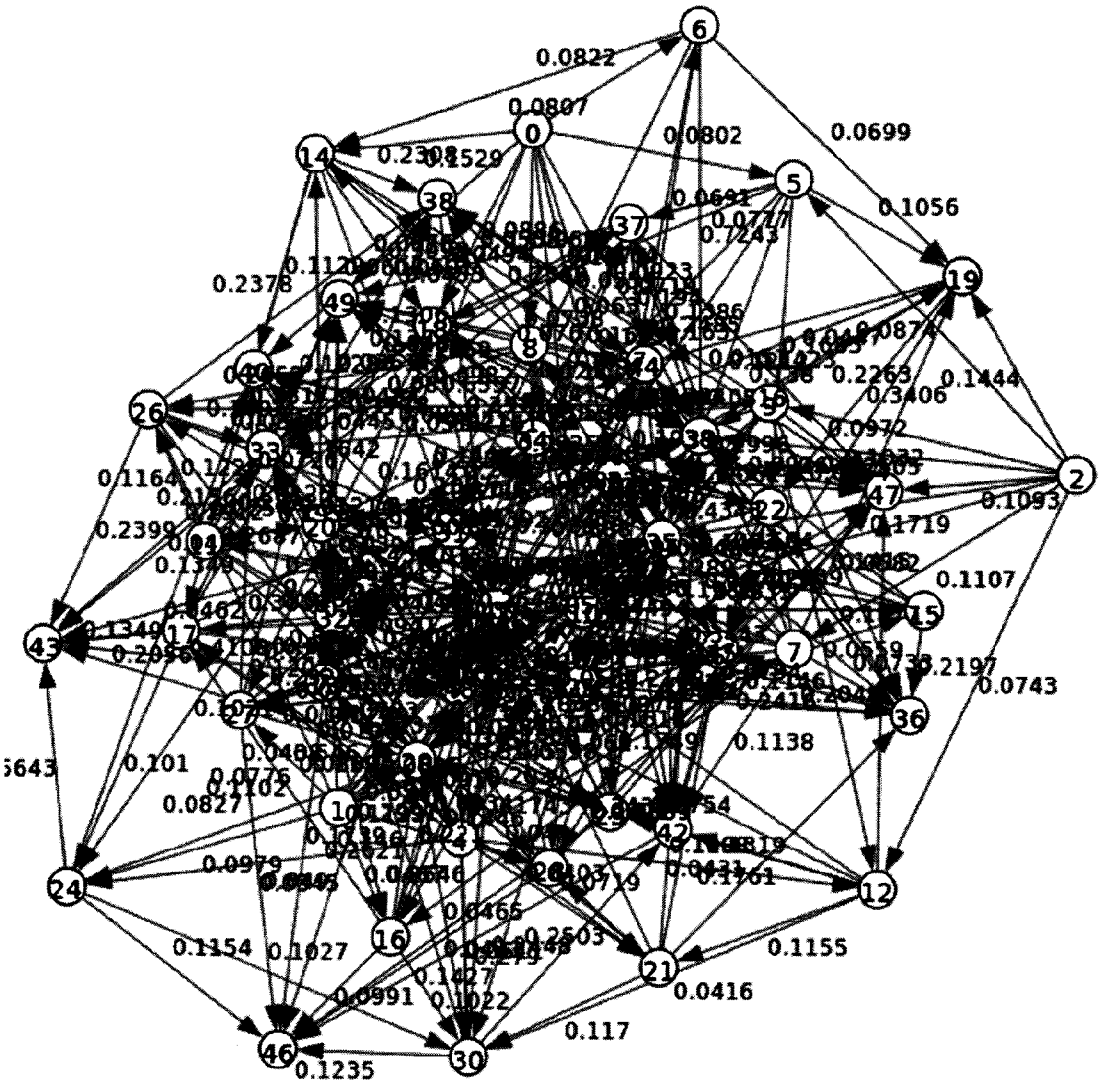
[図4]



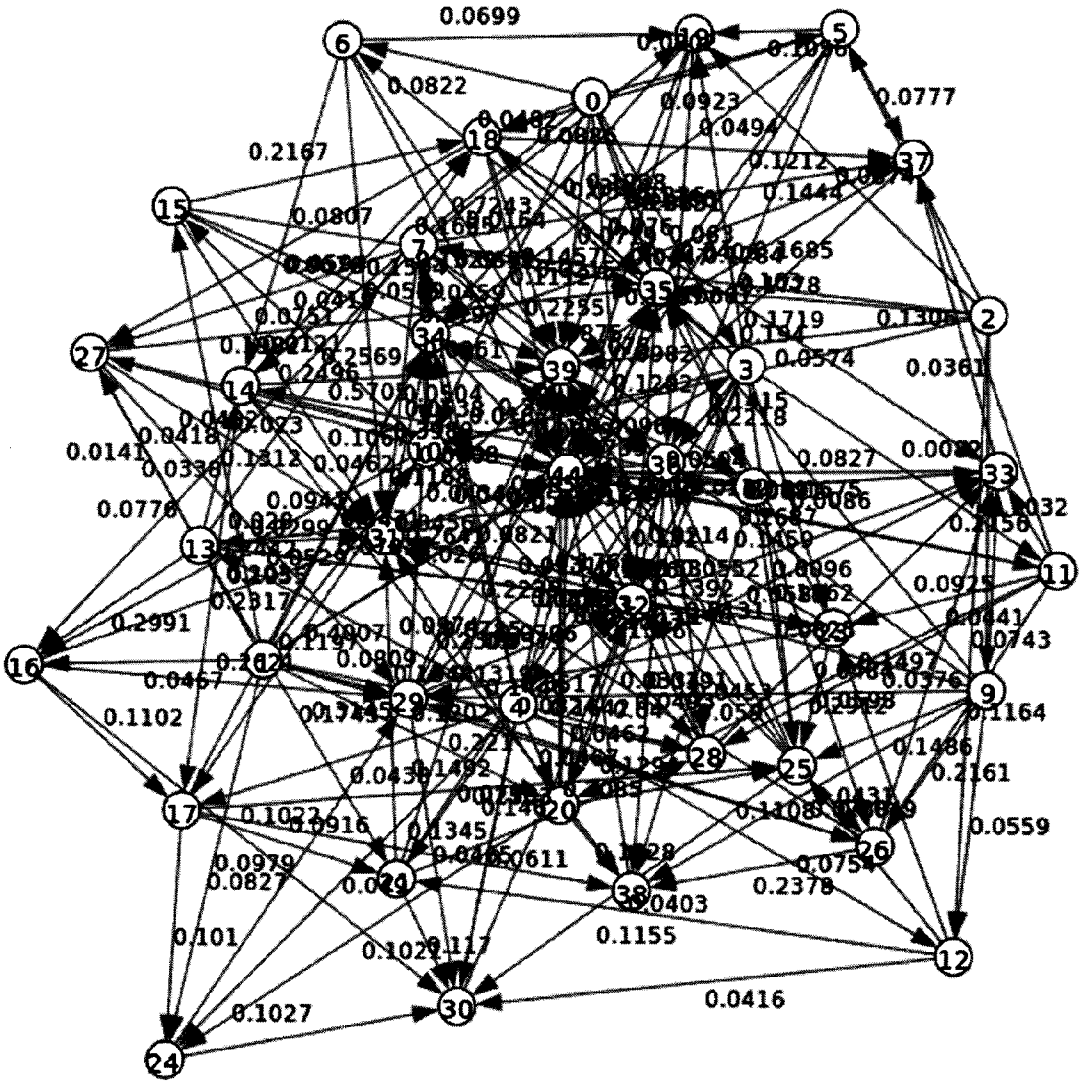
[図5]



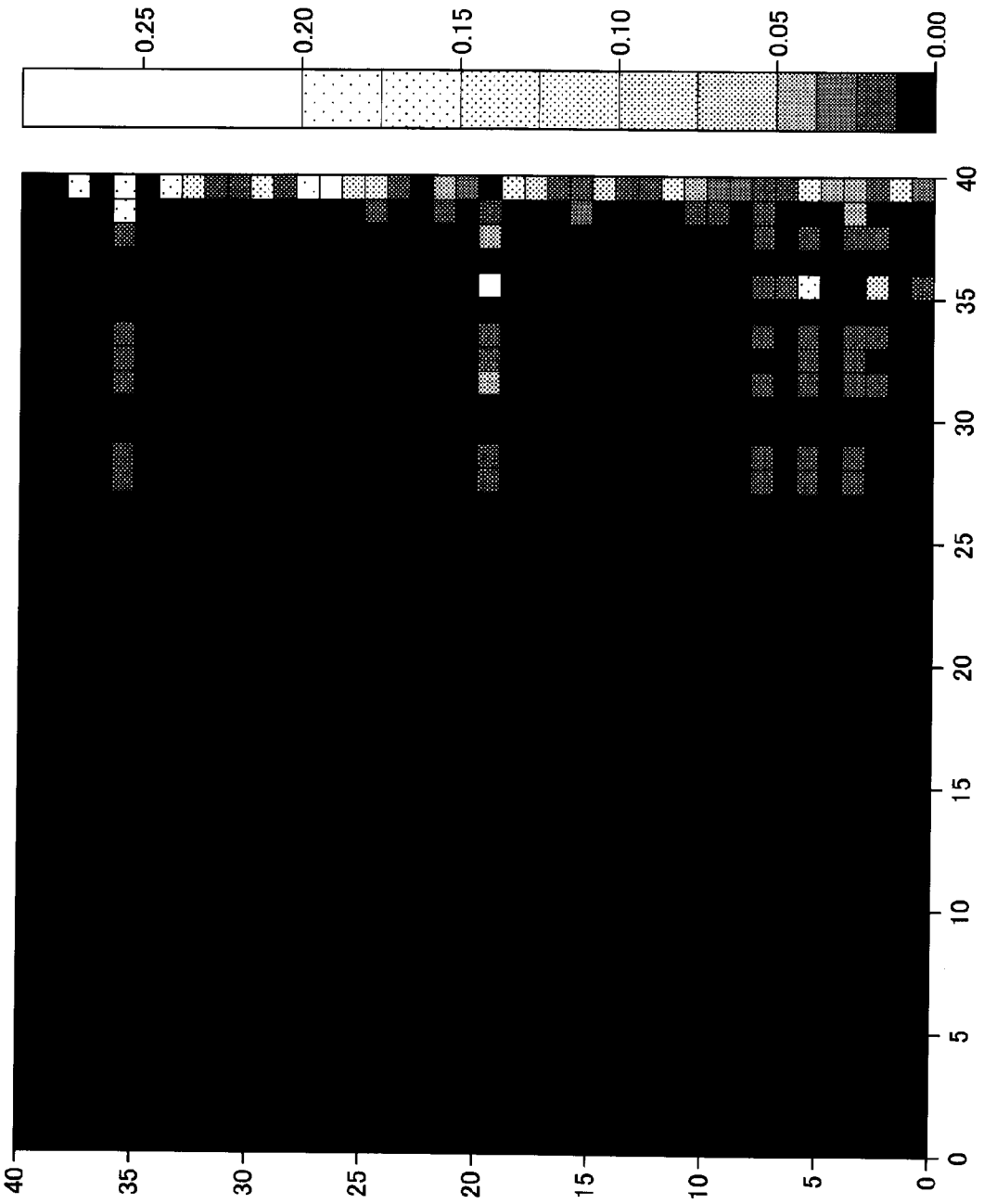
[図6]


6
0.0822
0.0807
0
0.0802
0.0699
14
0.2308
0.1529
5
38
0.0691
0.1056
37
0.0717
0.7243
19
0.2378
49
0.0874
8
0.2263
0.1444
0.3406
26
0.0972
2
0.1164
47
0.1093
0.2399
0.1719
0.1107
43
17
7
0.2197
0.0743
36
0.1138
0.101
0.0827
42
0.0979
24
12
0.1154
0.1027
16
0.2503
0.1155
21
0.0416
0.0991
0.1022
0.117
46
30
0.1235

[図7]


0.0699
0.0822
0.0777
0.0494
0.2167
0.1212
0.1444
0.0807
0.1719
0.1308
0.0574
0.0361
0.0827
0.0776
0.0086
0.1156
0.0096
0.0925
0.2991
0.0641
0.0743
0.0467
0.1102
0.1164
0.1486
0.2161
0.1022
0.0916
0.1345
0.0559
0.0754
0.0827
0.2378
0.0403
0.101
0.1155
0.1027
0.0416
0.1492
0.0611

[図8]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2019/000596

A. CLASSIFICATION OF SUBJECT MATTER
G06F 16/00(2019.01)i
FI: G06F16/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F16/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-124829 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 09.08.2018 (2018-08-09) entire text, all drawings	1-7
A	JP 2008-052494 A (SONY CORP.) 06.03.2008 (2008-03-06) entire text, all drawings	1-7
A	US 2018/0025040 A1 (EBAY INC.) 25.01.2018 (2018-01-25) entire text, all drawings	1-7

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 January 2020 (30.01.2020)	10 February 2020 (10.02.2020)

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2019/000596

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2018-124829 A	09 Aug. 2018	(Family: none)	
JP 2008-052494 A	06 Mar. 2008	US 2008/0052378 A1 entire text, all drawings	
US 2018/0025040 A1	25 Jan. 2018	US 2018/0349427 A1 entire text, all drawings WO 2018/017826 A1 entire text, all drawings	

国際調査報告

国際出願番号
PCT/JP2020/000596

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 16/00(2019.01)i
FI: G06F16/00

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G06F16/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２０年
日本国実用新案登録公報	１９９６－２０２０年
日本国登録実用新案公報	１９９４－２０２０年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	JP 2018-124829 A (日本電信電話株式会社) 09.08.2018 (2018-08-09) 全文,全図	1-7
A	JP 2008-052494 A (ソニー株式会社) 06.03.2008 (2008-03-06) 全文,全図	1-7
A	US 2018/0025040 A1 (EBAY INC.) 25.01.2018 (2018-01-25) 全文,全図	1-7

☐ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
“O” 口頭による開示、使用、展示等に言及する文献
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
“Y” 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
30.01.2020	10.02.2020

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

齊藤　貴孝 5N 4774

電話番号 03-3581-1101 内線 3586

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2020/000596

引用文献	公表日	パテントファミリー文献	公表日
JP 2018-124829 A	09.08.2018	（ファミリーなし）	
JP 2008-052494 A	06.03.2008	US 2008/0052378 A1 全文,全図	
US 2018/0025040 A1	25.01.2018	US 2018/0349427 A1 全文,全図 WO 2018/017826 A1 全文,全図	