

(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
2023年10月5日(05.10.2023)

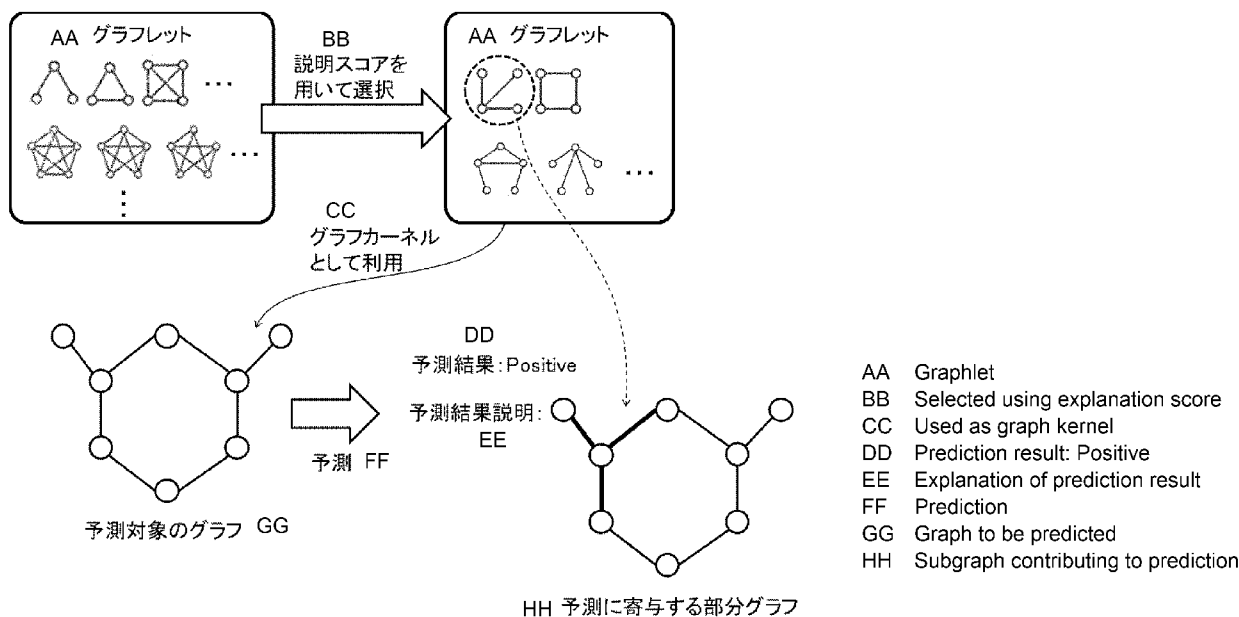


(10) 国際公開番号
WO 2023/188182 A1

- (51) 国際特許分類:
G06N 20/00 (2019.01)
- (21) 国際出願番号: PCT/JP2022/016245
- (22) 国際出願日: 2022年3月30日(30.03.2022)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者: 岡嶋 成司 (OKAJIMA, Seiji); 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP).
- (74) 代理人: 中島 淳, 外 (NAKAJIMA, Jun et al.); 〒1600022 東京都新宿区新宿4丁目3番17号 H K 新宿ビル7階 太陽国際特許事務所 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, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, 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,

(54) Title: SUBGRAPH STRUCTURE SELECTION PROGRAM, DEVICE, AND METHOD

(54) 発明の名称: 部分グラフ構造選択プログラム、装置、及び方法



(57) Abstract: This subgraph structure selection device calculates the occurrence rate of each graphlet in an explanation graph, calculates the explanation score of each graphlet on the basis of the contribution degree of each node which is added to the explanation graph, calculates, for each graphlet, the product of an average occurrence rate, a standard deviation of the occurrence rate, and an average explanation score in an explanation group set as an evaluation value, adds graphlets to a list in descending order of the evaluation value, calculates the accuracy of a prediction model for the case where an explanation graph having been vectorized utilizing the graphlets included in the list is adopted as input, and selects the graphlets added to the list as ultimate graphlets that are utilized in a graph kernel when a change of accuracy satisfies a

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, SC, 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))
-

prescribed condition.

(57) 要約 : 部分グラフ構造選択装置は、説明グラフにおける各グラフレットの出現頻度を算出し、説明グラフに付与された各ノードの寄与度に基づいて、各グラフレットの説明スコアを算出し、各グラフレットについて、説明グラフ集合における平均出現頻度と、出現頻度の標準偏差と、平均説明スコアとの積を評価値として算出し、評価値が大きい順にグラフレットをリストに追加し、リストに含まれるグラフレットを利用してベクトル化した説明グラフを入力とする場合の予測モデルの精度を算出し、精度の変化が所定条件を満たす場合に、リストに追加されているグラフレットを最終的にグラフカーネルで利用するグラフレットとして選択する。

明 細 書

発明の名称：部分グラフ構造選択プログラム、装置、及び方法
技術分野

[0001] 開示の技術は、部分グラフ構造選択プログラム、部分グラフ構造選択装置、及び部分グラフ構造選択方法に関する。

背景技術

[0002] ニューラルネットワーク等の機械学習モデルにグラフデータを入力し、タスクに応じた予測結果を得ると共に、入力されたグラフデータの中で、その予測結果に寄与した部分グラフを得る技術が存在する。この部分グラフは、機械学習モデルによる予測のプロセスを説明可能な情報である。また、このように予測のプロセスを説明可能な機械学習モデルを、説明可能AI（XAI：Explainable Artificial Intelligence）といい、入力データがグラフデータのXAIを、グラフXAIという。

[0003] グラフデータを機械学習モデルの入力とするために、グラフデータを高次元ベクトルに写像するグラフカーネルという技術が存在する。グラフカーネルとしては、例えば、Random walk kernel、Graphlet kernel、Weisfeiler-Lehman kernel等がある。これらのグラフカーネルは、多くの場合、写像されたベクトルの各要素は、プリミティブな部分グラフを示す。グラフXAIにおいては、できるだけ簡潔なグラフデータのベクトル表現が得られることが望ましい。

[0004] 例えば、Graphlet kernelは、少数のノードからなるグラフレットを列挙し、それぞれのグラフレットが、グラフ中に出現する回数を数え上げて、グラフをベクトル化する。グラフレットは、予め定めた数のノードを含み、ノード間の接続の全てのパターンを列挙したものである。ノード数が{3, 4, 5}の場合、グラフレットの数は29個となるため、ベクトルは29次元となる。このグラフレットを用いたグラフのベクトル化では、グラフレットの数え上げの計算コストがかかるという課題がある。計算コ

ストを抑えるために、例えば、グラフレットのノード数を {3, 4} にする等、ノード数を絞ることで、グラフレットの数を減らすことが考えられる。しかし、この場合、グラフデータのベクトルを用いた機械学習モデルの学習及び予測の精度に悪影響があるため、単純にグラフレットの数を減らすことはできない。

[0005] そこで、グラフのベクトル化の計算コストの削減と、学習及び予測の精度向上とのために、グラフレットの選択を行う技術が提案されている。この技術は、特定ドメインのグラフにおいて、多くの場合、特定のグラフレットの出現頻度は低くなることに着目し、グラフにおける出現頻度又は標準偏差が小さいグラフレットを削除する。また、この技術は、他のグラフレットと相関性の高い冗長なグラフレットを削除する。

先行技術文献

非特許文献

[0006] 非特許文献1: Furqan Aziz, Afan Ullah, Faiza Shah, "Feature selection and learning for graphlet kernel", Pattern Recognition Letters, Volume 136, p. 63-70, ISSN 0167-8655, August 2020.

発明の概要

発明が解決しようとする課題

[0007] しかしながら、上記の従来技術では、グラフ内において有意な部分グラフに相当するグラフレットが、低頻度又は冗長ということでグラフレットから削除されてしまう可能性がある。

[0008] 一つの側面として、開示の技術は、グラフカーネルに利用する部分グラフ構造として、有意な部分グラフ構造を選択することを目的とする。

課題を解決するための手段

[0009] 一つの態様として、開示の技術は、複数のノード及び複数のエッジを含む、1以上の予測対象のグラフのそれぞれにおける、予め定めた複数の部分グラフ構造のそれぞれの出現頻度を算出する。また、開示の技術は、前記1以

上の予測対象のグラフのそれぞれを、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する前記ノード又は前記エッジ毎の寄与度に基づいて、前記複数の部分グラフ構造のそれぞれの説明スコアを算出する。また、開示の技術は、前記複数の部分グラフ構造のそれぞれについて、前記1以上の予測対象のグラフにおける、前記出現頻度の平均と、前記出現頻度の標準偏差と、前記説明スコアの平均との積を算出する。また、開示の技術は、前記積が大きい順に、前記複数の部分グラフ構造の中から1つの部分グラフ構造を選択してリストに追加する。開示の技術は、その都度、前記リストに含まれる部分グラフ構造を利用してベクトル化した前記予測対象のグラフを入力とする場合の前記機械学習モデルの精度を算出する。そして、開示の技術は、前記精度の変化が所定条件を満たす場合に、前記リストに追加されている部分グラフ構造を最終的に利用する部分グラフ構造として選択する。

発明の効果

[0010] 一つの側面として、グラフカーネルに利用する部分グラフ構造として、有意な部分グラフ構造を選択することができる、という効果を有する。

図面の簡単な説明

[0011] [図1]部分グラフ構造選択装置の機能ブロック図である。

[図2]説明グラフの一例を示す図である。

[図3]グラフレットの選択の問題点を説明するための図である。

[図4]グラフレットの一例を示す図である。

[図5]グラフレットの出現頻度及び説明スコアの算出を説明するための図である。

[図6]部分グラフ構造選択装置として機能するコンピュータの概略構成を示すブロック図である。

[図7]部分グラフ構造選択処理の一例を示すフローチャートである。

[図8]選択されたグラフレットを用いた処理を説明するための図である。

発明を実施するための形態

[0012] 以下、図面を参照して、開示の技術に係る実施形態の一例を説明する。

- [0013] 図1に示すように、部分グラフ構造選択装置10には、説明グラフ集合が入力される。部分グラフ構造選択装置10は、説明グラフ集合に基づいて、グラフカーネルで利用するグラフレットを選択し、出力する。なお、グラフレットは、開示の技術の「部分グラフ構造」の一例である。
- [0014] 説明グラフとは、複数のノードと、ノード間を接続する複数のエッジとを含むグラフであって、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する寄与度、すなわち予測への貢献度が、ノード又はエッジ毎に付与されたグラフである。本実施形態では、寄与度が各ノードに付与される場合を例として説明する。
- [0015] 図2上図に、説明グラフの一例を示す。図2上図では、化学構造を表すグラフの例を示している。各ノード（丸印）に併記した数字が寄与度である。本実施形態では、予測結果に対して有意な部分に含まれるノードの寄与度の平均は高くなるとの仮定の下、グラフレットの選択に寄与度を利用する。図2のような化学構造の場合、この仮定は、化学的に有意な構造を構成するノードの寄与度の平均が高くなることに相当する。例えば、図2上図の説明グラフにおいて、図2下図のAに示す部分グラフ（以下、「部分グラフA」という）とBに示す部分グラフ（以下、「部分グラフB」という）とでは、部分グラフBの方が寄与度の平均は高くなる。したがって、部分グラフBの方が、説明グラフにおいて有意な構造であることを表している。
- [0016] ここで、上述した従来技術のように、グラフカーネルに利用するグラフレットを選択する際に、説明グラフにおける出現頻度又は標準偏差が小さいグラフレットを削除すると共に、他のグラフレットと相関性の高い冗長なグラフレットを削除とする。上記の図2の例において、部分グラフAと部分グラフBとが高頻度で同時に出現する場合、図3に示すように、冗長であるという理由で、化学的に意味がありそうな構造C_H₃を表す部分グラフと構造が一致するグラフレットが削除される可能性がある。すなわち、グラフカーネルに利用するグラフレットから、有意なグラフレットが除外される可能性がある。そこで、本実施形態では、上述したように、説明グラフの各ノード

に付与された寄与度を、グラフレットの選択に利用するものである。

- [0017] 部分グラフ構造選択装置10は、機能的には、出現頻度算出部12と、説明スコア算出部14と、評価値算出部16と、選択部18と、削除部20とを含む。また、部分グラフ構造選択装置10の所定の記憶領域には、訓練済みの機械学習モデルである予測モデル30が記憶される。なお、評価値算出部16は、開示の技術の「積算出部」の一例である。
- [0018] 出現頻度算出部12は、説明グラフ集合に含まれる説明グラフのそれぞれにおける、予め定めた複数のグラフレットのそれぞれの出現頻度を算出する。予め定めた複数のグラフレットとしては、例えば、図4に示すように、ノード数が{3, 4, 5}の、 $g_1 \sim g_{29}$ の29個のグラフレットを定めておいてよい。図4及び後述する図5及び図8の一部に用いたグラフレットの図は、非特許文献1の図面を引用したものである。出現頻度算出部12は、例えば図5に示すように、各グラフレット（図5の例では、グラフレット g_6 ）と構造が一致する部分グラフを説明グラフから探索してカウントすることで、出現頻度を算出する。図5の例では、グラフレット g_6 と構造が一致する部分グラフとして、部分グラフA（破線で示す部分グラフ）と、部分グラフB（一点鎖線で示す部分グラフ）とが存在する。したがって、出現頻度算出部12は、グラフレット g_6 の出現頻度を「2」と算出する。
- [0019] 説明スコア算出部14は、説明グラフのノード毎の寄与度に基づいて、各グラフレットの説明スコアを算出する。具体的には、説明スコア算出部14は、説明グラフの中で、グラフレットの構造と一致する部分グラフに含まれるノードの寄与度の平均を、そのグラフレットの説明スコアとして算出する。
- [0020] 1つの説明グラフの中に、グラフレットの構造と一致する部分グラフが複数ある場合、説明スコア算出部14は、複数の部分グラフのそれぞれについて算出した説明スコアのうち高い方を、そのグラフレットの説明スコアとする。図5の例の場合、部分グラフAの説明スコアは、 $(0.2 + 0.6 + 0.9 + 0.1) / 4 = 0.45$ 、部分グラフBの説明スコアは、 $(0.8 +$

$0.9 + 0.7 + 0.7) / 4 = 0.775$ である。したがって、説明スコア算出部14は、該当の説明グラフに対するグラフレット g_6 の説明スコアを 0.775 と算出する。

[0021] なお、説明スコア算出部14は、グラフレットの構造と一致する部分グラフが複数ある場合に、説明スコアの高い方を選択する場合に限らず、複数の部分グラフについての説明スコアの平均を、該当のグラフレットの説明スコアとして算出してもよい。

[0022] 評価値算出部16は、複数のグラフレットのそれぞれについて、説明グラフ集合における、出現頻度の平均と、出現頻度の標準偏差と、説明スコアの平均との積を評価値として算出する。具体的には、評価値算出部16は、グラフレット g_i について、各説明グラフから算出された出現頻度の全説明グラフでの平均（以下、「平均出現頻度」という） μ_i を算出する。また、評価値算出部16は、グラフレット g_i について、各説明グラフから算出された出現頻度の全説明グラフにおける標準偏差 σ_i を算出する。また、評価値算出部16は、グラフレット g_i について、各説明グラフから算出された説明スコアの全説明グラフにおける平均（以下、「平均説明スコア」という） s_i を算出する。そして評価値算出部16は、平均出現頻度 μ_i と標準偏差 σ_i と平均説明スコア s_i との積を、グラフレット g_i の評価値 $\mu \sigma s_i$ として算出する。

[0023] 選択部18は、評価値算出部16により算出された評価値が大きい順に、複数のグラフレットの中から1つのグラフレットを選択してリストに追加する。選択部18は、その都度、リストに含まれるグラフレットを利用してベクトル化した説明グラフを入力とする場合の予測モデル30の精度を算出する。選択部18は、精度の変化が所定条件を満たす場合に、リストを削除部20へ受け渡す。選択部18は、所定条件を、精度が上がらなくなった場合、又は低下する場合としてよい。選択部18は、前回算出された精度と今回算出された精度との差が所定値以内となった場合を、精度が上がらなくなった場合と判定してよい。また、選択部18は、前回算出された精度より今回算出された精度の方が低い場合を、精度が低下する場合と判定してよい。

- [0024] 削除部20は、リストに追加されているグラフレットについての全てのペアの相関を示す指標を算出し、指標が所定値以上のペアについて、平均説明スコアsが低い方のグラフレットをリストから削除する。削除部20は、相関を示す指標として相互相関cを算出してよい。相関性の高いグラフレット同士を両方残した場合には冗長となるため、一方を削除するものである。その際に、平均説明スコアsが低い方のグラフレットを削除することで、有意な構造のグラフレットが残り易くなる。削除部20は、リストに残っているグラフレットを最終的にグラフカーネルで利用するグラフレットとして出力する。
- [0025] 部分グラフ構造選択装置10は、例えば図6に示すコンピュータ40で実現されてよい。コンピュータ40は、CPU (Central Processing Unit) 41と、一時記憶領域としてのメモリ42と、不揮発性の記憶部43とを備える。また、コンピュータ40は、入力部、表示部等の入出力装置44と、記憶媒体49に対するデータの読み込み及び書き込みを制御するR/W (Read/Write) 部45とを備える。また、コンピュータ40は、インターネット等のネットワークに接続される通信I/F (Interface) 46を備える。CPU 41、メモリ42、記憶部43、入出力装置44、R/W部45、及び通信I/F 46は、バス47を介して互いに接続される。
- [0026] 記憶部43は、HDD (Hard Disk Drive)、SSD (Solid State Drive)、フラッシュメモリ等によって実現されてよい。記憶媒体としての記憶部43には、コンピュータ40を、部分グラフ構造選択装置10として機能させるための部分グラフ構造選択プログラム50が記憶される。部分グラフ構造選択プログラム50は、出現頻度算出プロセス52と、説明スコア算出プロセス54と、評価値算出プロセス56と、選択プロセス58と、削除プロセス60とを有する。また、記憶部43は、予測モデル30を構成する情報が記憶される情報記憶領域70を有する。
- [0027] CPU 41は、部分グラフ構造選択プログラム50を記憶部43から読み出してメモリ42に展開し、部分グラフ構造選択プログラム50が有する8

ロセスを順次実行する。CPU 41は、出現頻度算出プロセス52を実行することで、図1に示す出現頻度算出部12として動作する。また、CPU 41は、説明スコア算出プロセス54を実行することで、図1に示す説明スコア算出部14として動作する。また、CPU 41は、評価値算出プロセス56を実行することで、図1に示す評価値算出部16として動作する。また、CPU 41は、選択プロセス58を実行することで、図1に示す選択部18として動作する。また、CPU 41は、削除プロセス60を実行することで、図1に示す削除部20として動作する。また、CPU 41は、情報記憶領域70から情報を読み出して、予測モデル30をメモリ42に展開する。これにより、部分グラフ構造選択プログラム50を実行したコンピュータ40が、部分グラフ構造選択装置10として機能することになる。なお、プログラムを実行するCPU 41はハードウェアである。

[0028] なお、部分グラフ構造選択プログラム50により実現される機能は、例えば半導体集積回路、より詳しくはASIC (Application Specific Integrated Circuit) 等で実現することも可能である。

[0029] 次に、本実施形態に係る部分グラフ構造選択装置10の作用について説明する。部分グラフ構造選択装置10に説明グラフ集合が入力され、グラフレットの選択が指示されると、部分グラフ構造選択装置10において、図7に示す部分グラフ構造選択処理が実行される。なお、部分グラフ構造選択処理は、開示の技術の部分グラフ構造選択方法の一例である。

[0030] ステップS10で、出現頻度算出部12が、部分グラフ構造選択装置10に入力された説明グラフ集合を取得する。次に、ステップS12で、出現頻度算出部12が、グラフレットと構造が一致する部分グラフを説明グラフから探索してカウントすることにより、各説明グラフにおける、各グラフレットの出現頻度を算出する。

[0031] 次に、ステップS14で、説明スコア算出部14が、説明グラフの中で、グラフレットの構造と一致する部分グラフに含まれるノードの寄与度の平均を、そのグラフレットの説明スコアとして算出する。説明スコア算出部14

は、各説明グラフにおける、各グラフレットの説明スコアを算出する。

[0032] 次に、ステップS 16で、評価値算出部16が、各グラフレットについて、各説明グラフから算出された出現頻度の平均である平均出現頻度と、出現頻度の標準偏差と、各説明グラフから算出された説明スコアの平均である平均説明スコアとを算出する。そして評価値算出部16が、平均出現頻度と標準偏差と平均説明スコアとの積を、各グラフレットの評価値として算出する。

[0033] 次に、ステップS 18で、選択部18が、上記ステップS 16で算出された評価値が大きい順に、複数のグラフレットをソートしたリストLを作成する。次に、ステップS 20で、選択部18が、リストLから評価値最大のグラフレットを選択してリストL'に追加すると共に、リストLから削除する。

[0034] 次に、ステップS 22で、選択部18が、リストL'に含まれるグラフレットをグラフカーネルとして利用してベクトル化した説明グラフを入力とする場合の予測モデル30の精度を算出する。次に、ステップS 24で、選択部18が、上記ステップS 22で算出した精度が、前回算出された精度よりも低下したか否かを判定する。精度が低下していない場合には、ステップS 20に戻り、低下した場合には、ステップS 26へ移行する。

[0035] ステップS 26では、選択部18が、リストL'に最後に追加したグラフレットをリストL'から削除し、リストL'を削除部20へ受け渡す。次に、ステップS 28で、削除部20が、リストL'内のグラフレットについての全てのペアの相関を示す指標を算出する。そして、削除部20が、相関を示す指標が所定値以上のペアについて、平均説明スコアsが低い方のグラフレットをリストL'から削除する。削除部20は、リストL'に残っているグラフレットを最終的にグラフカーネルで利用するグラフレットとして出力し、部分グラフ構造選択処理を終了する。

[0036] 以上説明したように、本実施形態に係る部分グラフ構造選択装置は、複数のノード及び複数のエッジを含む、1以上の説明グラフのそれぞれにおける

、予め定めた複数のグラフレットのそれぞれの出現頻度を算出する。また、部分グラフ構造選択装置は、説明グラフに付与された各ノードの寄与度に基づいて、複数のグラフレットのそれぞれの説明スコアを算出する。また、部分グラフ構造選択装置は、複数の部分グラフレットのそれぞれについて、説明グラフ集合における、平均出現頻度、出現頻度の標準偏差と、平均説明スコアとの積を評価値として算出する。また、部分グラフ構造選択装置は、評価値が大きい順に、複数のグラフレットの中から1つのグラフレットを選択してリストに追加する。部分グラフ構造選択装置は、その都度、リストに含まれるグラフレットを利用してベクトル化した説明グラフを入力とする場合の予測モデルの精度を算出する。そして、部分グラフ構造選択装置は、精度の変化が所定条件を満たす場合に、リストに追加されているグラフレットを最終的にグラフカーネルで利用する部分グラフ構造として選択する。これにより、グラフカーネルに利用する部分グラフ構造として、有意な部分グラフ構造を選択することができる。

[0037] 例えば、図8に示すように、本実施形態は、グラフカーネルに利用するグラフレットを、説明グラフにおける寄与度に基づく説明スコアを用いて選択することにより、有意な部分グラフ構造を失うことなく、簡潔な組み合わせのグラフレットを選択することができる。そして、このように選択されたグラフレットをグラフカーネルとして利用して、予測対象のグラフに対する機械学習モデルによる予測を行うことで、予測結果と共に得られる予測結果説明も有意性を失うことなく、簡潔な組み合わせで表現することができる。図8の例では、選択されたグラフレットのうち、破線で囲ったグラフレットと構造が一致する部分グラフ（太線部分）が、予測に寄与する部分グラフとして特定された例を示している。

[0038] その結果、予測結果及び予測結果説明に基づいて後段の因果推論等を行う際に、グラフ内の部分グラフ間の因果関係等として、有意な因果関係が推定され易くなる。例えば、化学構造を表すグラフの場合、因果推論を行うことで、反応メカニズムに関わる部分グラフを発見することに寄与し得る。

[0039] なお、上記実施形態では、部分グラフ構造選択プログラムが記憶部に予め記憶（インストール）されている様子を説明したが、これに限定されない。開示の技術に係るプログラムは、CD-ROM、DVD-ROM、USBメモリ等の記憶媒体に記憶された形態で提供することも可能である。

符号の説明

- [0040]
- 10 部分グラフ構造選択装置
 - 12 出現頻度算出部
 - 14 説明スコア算出部
 - 16 評価値算出部
 - 18 選択部
 - 20 削除部
 - 30 予測モデル
 - 40 コンピュータ
 - 41 CPU
 - 42 メモリ
 - 43 記憶部
 - 44 入出力装置
 - 45 R/W部
 - 46 通信I/F
 - 47 バス
 - 49 記憶媒体
 - 50 部分グラフ構造選択プログラム
 - 52 出現頻度算出プロセス
 - 54 説明スコア算出プロセス
 - 56 評価値算出プロセス
 - 58 選択プロセス
 - 60 削除プロセス
 - 70 情報記憶領域

請求の範囲

- [請求項1] 複数のノード及び複数のエッジを含む、1以上の予測対象のグラフのそれぞれにおける、予め定めた複数の部分グラフ構造のそれぞれの出現頻度を算出し、
- 前記1以上の予測対象のグラフのそれぞれを、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する前記ノード又は前記エッジ毎の寄与度に基づいて、前記複数の部分グラフ構造のそれぞれの説明スコアを算出し、
- 前記複数の部分グラフ構造のそれぞれについて、前記1以上の予測対象のグラフにおける、前記出現頻度の平均と、前記出現頻度の標準偏差と、前記説明スコアの平均との積を算出し、
- 前記積が大きい順に、前記複数の部分グラフ構造の中から1つの部分グラフ構造を選択してリストに追加する都度、前記リストに含まれる部分グラフ構造を利用してベクトル化した前記予測対象のグラフを入力とする場合の前記機械学習モデルの精度を算出し、前記精度の変化が所定条件を満たす場合に、前記リストに追加されている部分グラフ構造を最終的に利用する部分グラフ構造として選択することを含む処理をコンピュータに実行させるための部分グラフ構造選択プログラム。
- [請求項2] 前記リストに追加されている部分グラフ構造についての全てのペアの相関を示す指標を算出し、前記指標が所定値以上のペアについて、前記説明スコアが低い方の部分グラフ構造を前記リストから削除する請求項1に記載の部分グラフ構造選択プログラム。
- [請求項3] 前記説明スコアとして、1つの前記予測対象のグラフの中で、前記部分グラフ構造と構造が一致する部分グラフに含まれる前記ノード又は前記エッジの前記寄与度の平均を算出する請求項1又は請求項2に記載の部分グラフ構造選択プログラム。
- [請求項4] 1つの前記予測対象のグラフの中に、前記部分グラフ構造と構造が

一致する部分グラフが複数ある場合、複数の部分グラフのそれぞれについて算出された前記寄与度の平均のうち高い方の値を前記部分グラフ構造の説明スコアとする請求項3に記載の部分グラフ構造選択プログラム。

[請求項5] 前記所定条件を、前回算出された精度と今回算出された精度との差が所定値以内となった場合、又は前回算出された精度より今回算出された精度の方が低い場合とする請求項1又は請求項2に記載の部分グラフ構造選択プログラム。

[請求項6] 複数のノード及び複数のエッジを含む、1以上の予測対象のグラフのそれぞれにおける、予め定めた複数の部分グラフ構造のそれぞれの出現頻度を算出する出現頻度算出部と、

前記1以上の予測対象のグラフのそれぞれを、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する前記ノード又は前記エッジ毎の寄与度に基づいて、前記複数の部分グラフ構造のそれぞれの説明スコアを算出する説明スコア算出部と、

前記複数の部分グラフ構造のそれぞれについて、前記1以上の予測対象のグラフにおける、前記出現頻度の平均と、前記出現頻度の標準偏差と、前記説明スコアの平均との積を算出する積算出部と、

前記積が大きい順に、前記複数の部分グラフ構造の中から1つの部分グラフ構造を選択してリストに追加する都度、前記リストに含まれる部分グラフ構造を利用してベクトル化した前記予測対象のグラフを入力とする場合の前記機械学習モデルの精度を算出し、前記精度の変化が所定条件を満たす場合に、前記リストに追加されている部分グラフ構造を最終的に利用する部分グラフ構造として選択する選択部と、
を含む部分グラフ構造選択装置。

[請求項7] 前記リストに追加されている部分グラフ構造についての全てのペアの相関を示す指標を算出し、前記指標が所定値以上のペアについて、前記説明スコアが低い方の部分グラフ構造を前記リストから削除する

削除部をさらに含む請求項6に記載の部分グラフ構造選択装置。

[請求項8] 前記説明スコア算出部は、前記説明スコアとして、1つの前記予測対象のグラフの中で、前記部分グラフ構造と構造が一致する部分グラフに含まれる前記ノード又は前記エッジの前記寄与度の平均を算出する請求項6又は請求項7に記載の部分グラフ構造選択装置。

[請求項9] 前記説明スコア算出部は、1つの前記予測対象のグラフの中に、前記部分グラフ構造と構造が一致する部分グラフが複数ある場合、複数の部分グラフのそれぞれについて算出された前記寄与度の平均のうち高い方の値を前記部分グラフ構造の説明スコアとする請求項8に記載の部分グラフ構造選択装置。

[請求項10] 前記所定条件を、前回算出された精度と今回算出された精度との差が所定値以内となった場合、又は前回算出された精度より今回算出された精度の方が低い場合とする請求項6又は請求項7に記載の部分グラフ構造選択装置。

[請求項11] 複数のノード及び複数のエッジを含む、1以上の予測対象のグラフのそれぞれにおける、予め定めた複数の部分グラフ構造のそれぞれの出現頻度を算出し、

前記1以上の予測対象のグラフのそれぞれを、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する前記ノード又は前記エッジ毎の寄与度に基づいて、前記複数の部分グラフ構造のそれぞれの説明スコアを算出し、

前記複数の部分グラフ構造のそれぞれについて、前記1以上の予測対象のグラフにおける、前記出現頻度の平均と、前記出現頻度の標準偏差と、前記説明スコアの平均との積を算出し、

前記積が大きい順に、前記複数の部分グラフ構造の中から1つの部分グラフ構造を選択してリストに追加する都度、前記リストに含まれる部分グラフ構造を利用してベクトル化した前記予測対象のグラフを入力とする場合の前記機械学習モデルの精度を算出し、前記精度の変

化が所定条件を満たす場合に、前記リストに追加されている部分グラフ構造を最終的に利用する部分グラフ構造として選択する

ことを含む処理をコンピュータが実行する部分グラフ構造選択方法

。

[請求項12] 前記リストに追加されている部分グラフ構造についての全てのペアの相関を示す指標を算出し、前記指標が所定値以上のペアについて、前記説明スコアが低い方の部分グラフ構造を前記リストから削除する請求項11に記載の部分グラフ構造選択方法。

[請求項13] 前記説明スコアとして、1つの前記予測対象のグラフの中で、前記部分グラフ構造と構造が一致する部分グラフに含まれる前記ノード又は前記エッジの前記寄与度の平均を算出する請求項11又は請求項12に記載の部分グラフ構造選択方法。

[請求項14] 1つの前記予測対象のグラフの中に、前記部分グラフ構造と構造が一致する部分グラフが複数ある場合、複数の部分グラフのそれぞれについて算出された前記寄与度の平均のうち高い方の値を前記部分グラフ構造の説明スコアとする請求項13に記載の部分グラフ構造選択方法。

[請求項15] 前記所定条件を、前回算出された精度と今回算出された精度との差が所定値以内となった場合、又は前回算出された精度より今回算出された精度の方が低い場合とする請求項11又は請求項12に記載の部分グラフ構造選択方法。

[請求項16] 複数のノード及び複数のエッジを含む、1以上の予測対象のグラフのそれぞれにおける、予め定めた複数の部分グラフ構造のそれぞれの出現頻度を算出し、

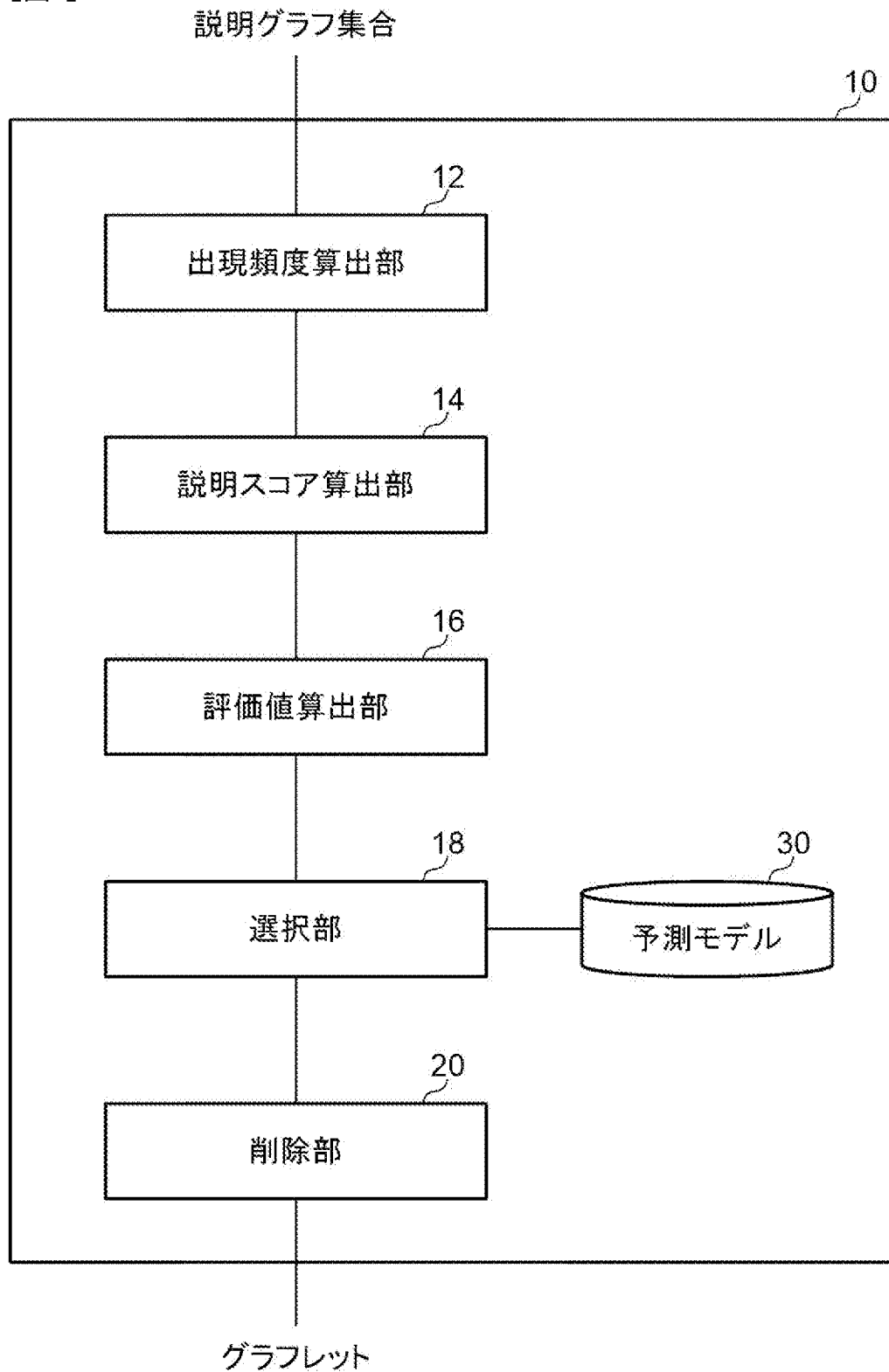
前記1以上の予測対象のグラフのそれぞれを、訓練済みの機械学習モデルに入力した際に出力される予測結果に対する前記ノード又は前記エッジ毎の寄与度に基づいて、前記複数の部分グラフ構造のそれぞれの説明スコアを算出し、

前記複数の部分グラフ構造のそれぞれについて、前記1以上の予測対象のグラフにおける、前記出現頻度の平均と、前記出現頻度の標準偏差と、前記説明スコアの平均との積を算出し、

前記積が大きい順に、前記複数の部分グラフ構造の中から1つの部分グラフ構造を選択してリストに追加する都度、前記リストに含まれる部分グラフ構造を利用してベクトル化した前記予測対象のグラフを入力とする場合の前記機械学習モデルの精度を算出し、前記精度の変化が所定条件を満たす場合に、前記リストに追加されている部分グラフ構造を最終的に利用する部分グラフ構造として選択する

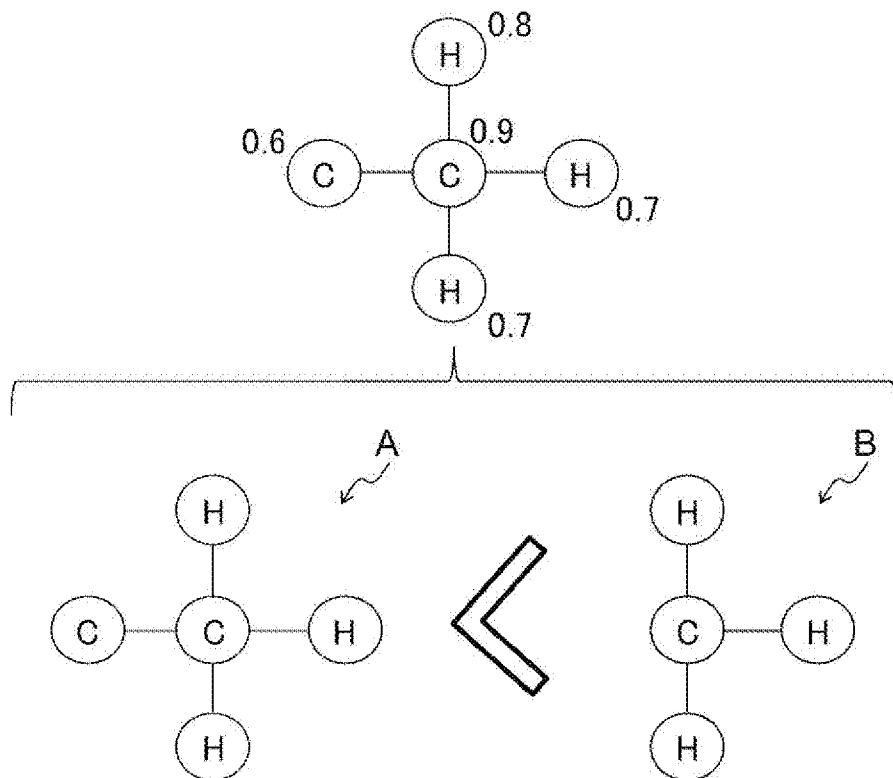
ことを含む処理をコンピュータに実行させるための部分グラフ構造選択プログラムを記憶した非一時的記憶媒体。

[図1]

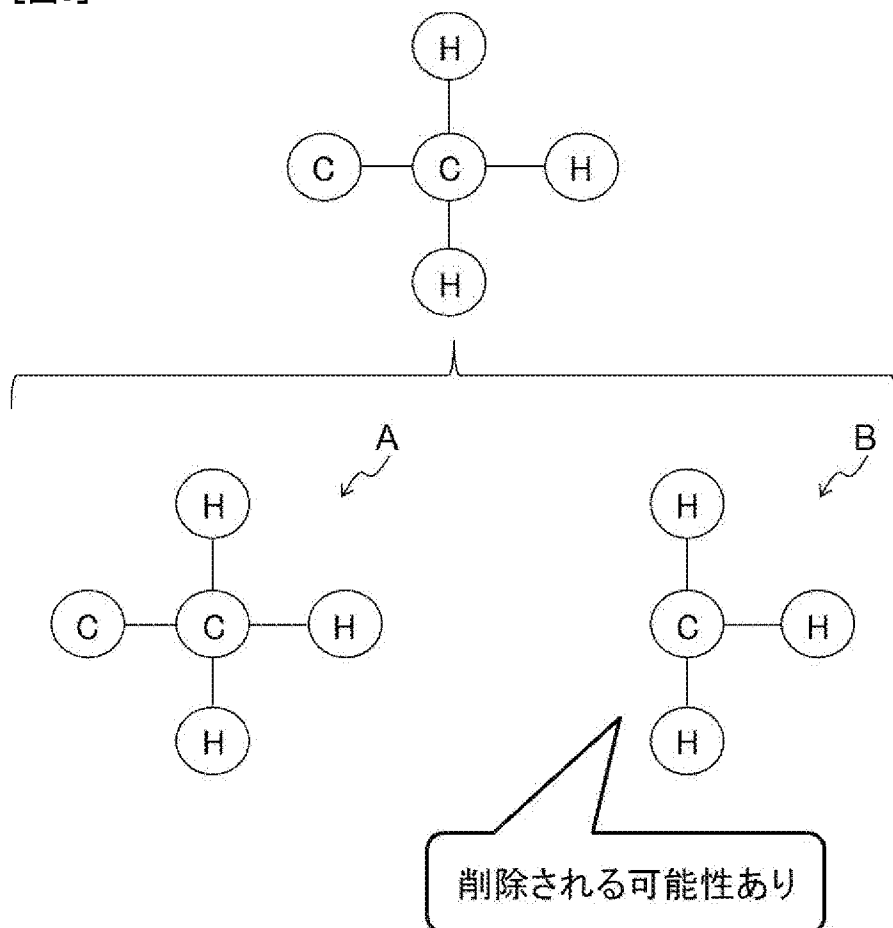


[図2]

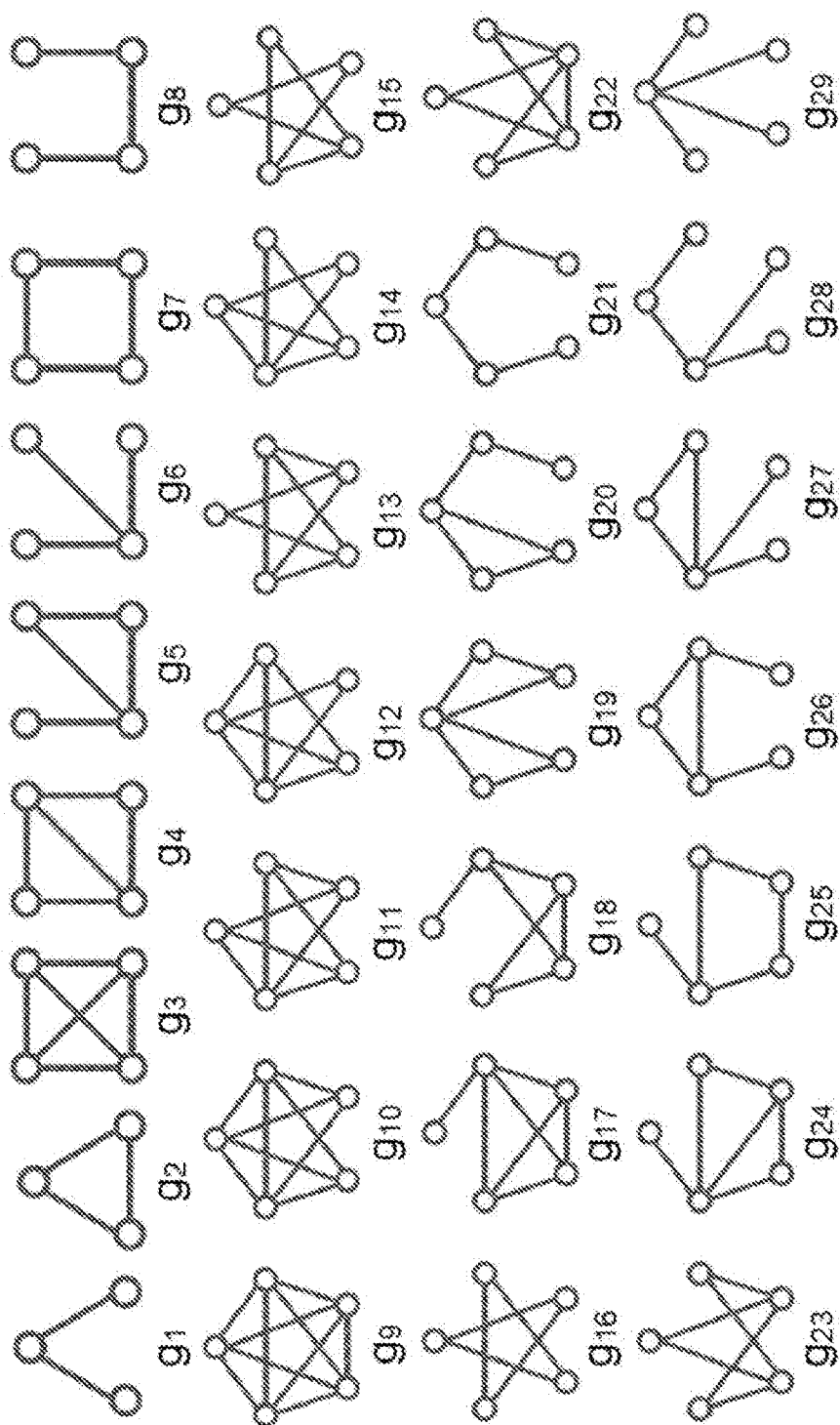
説明グラフ



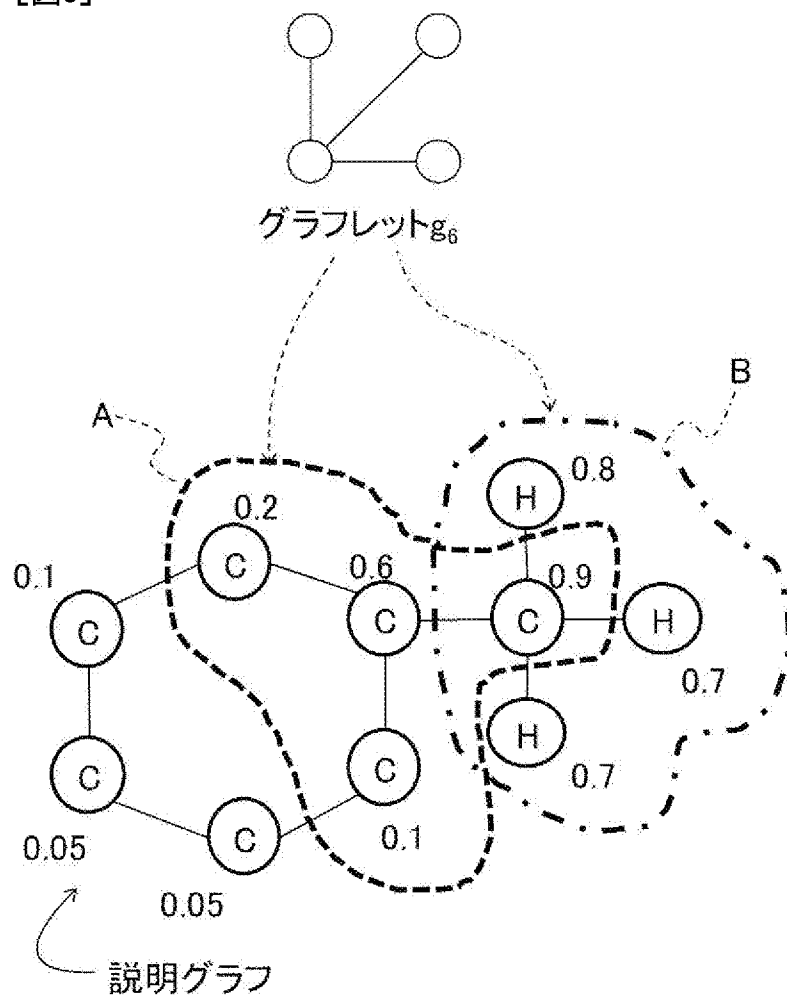
[図3]



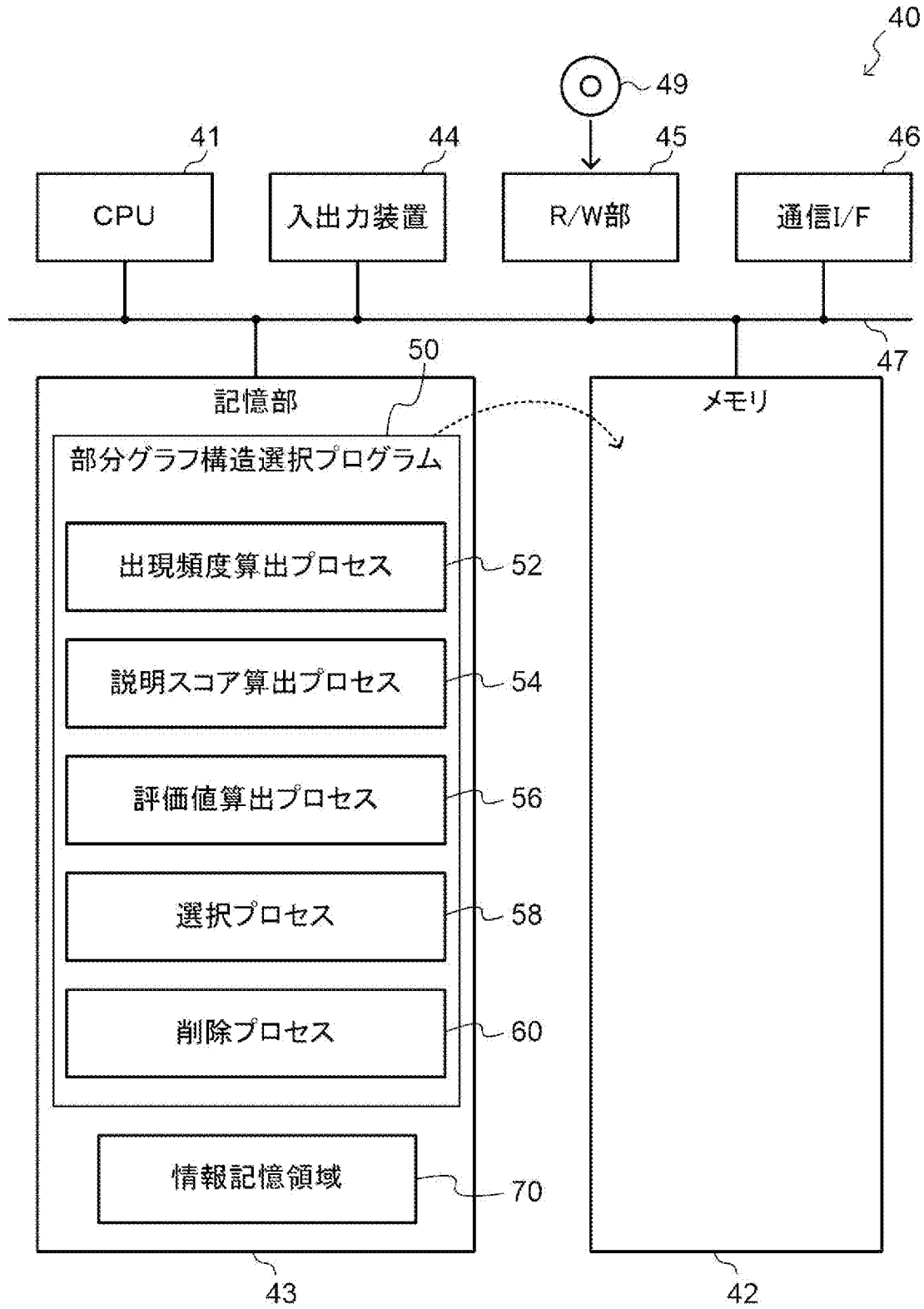
[図4]



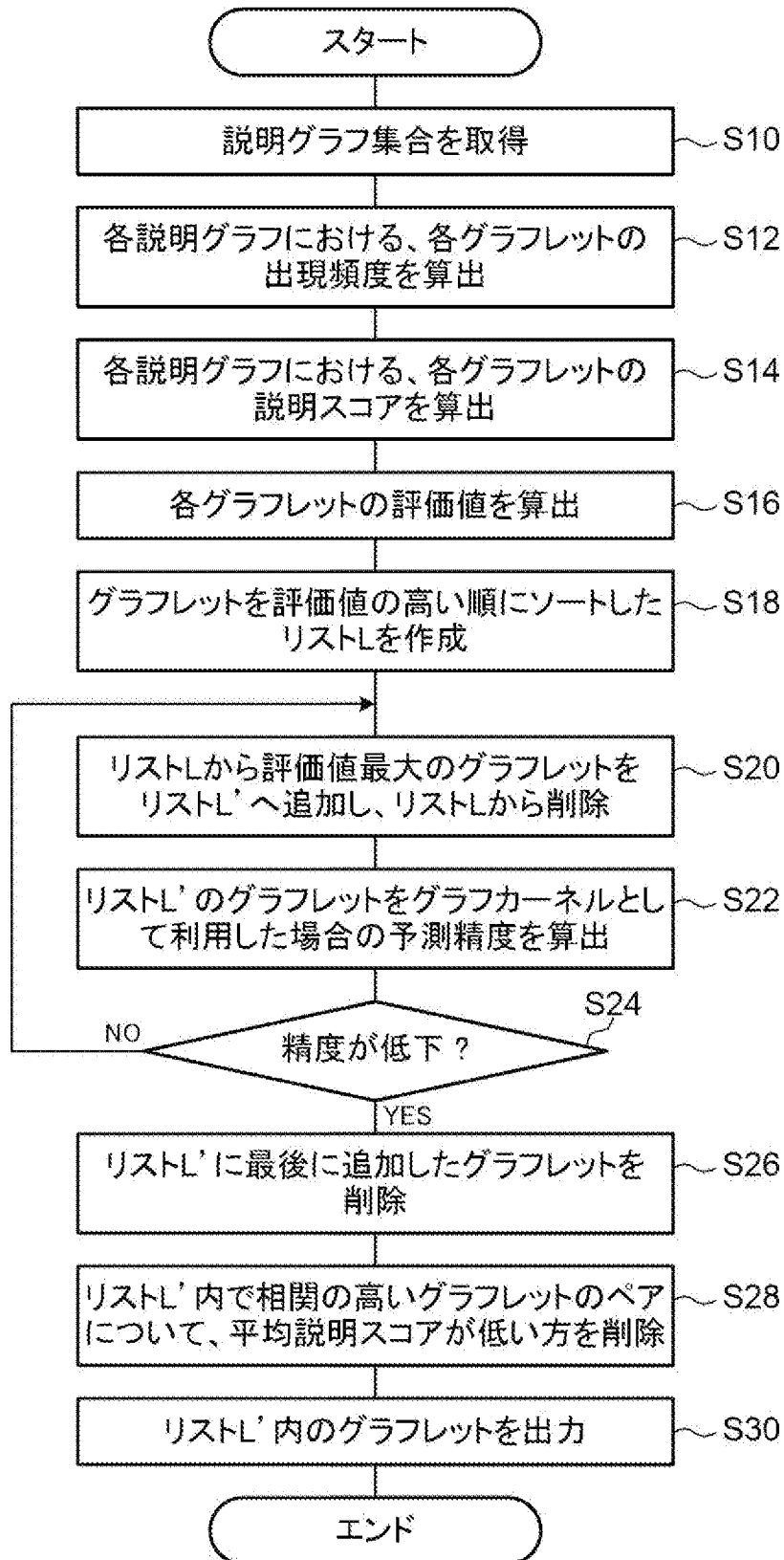
[図5]



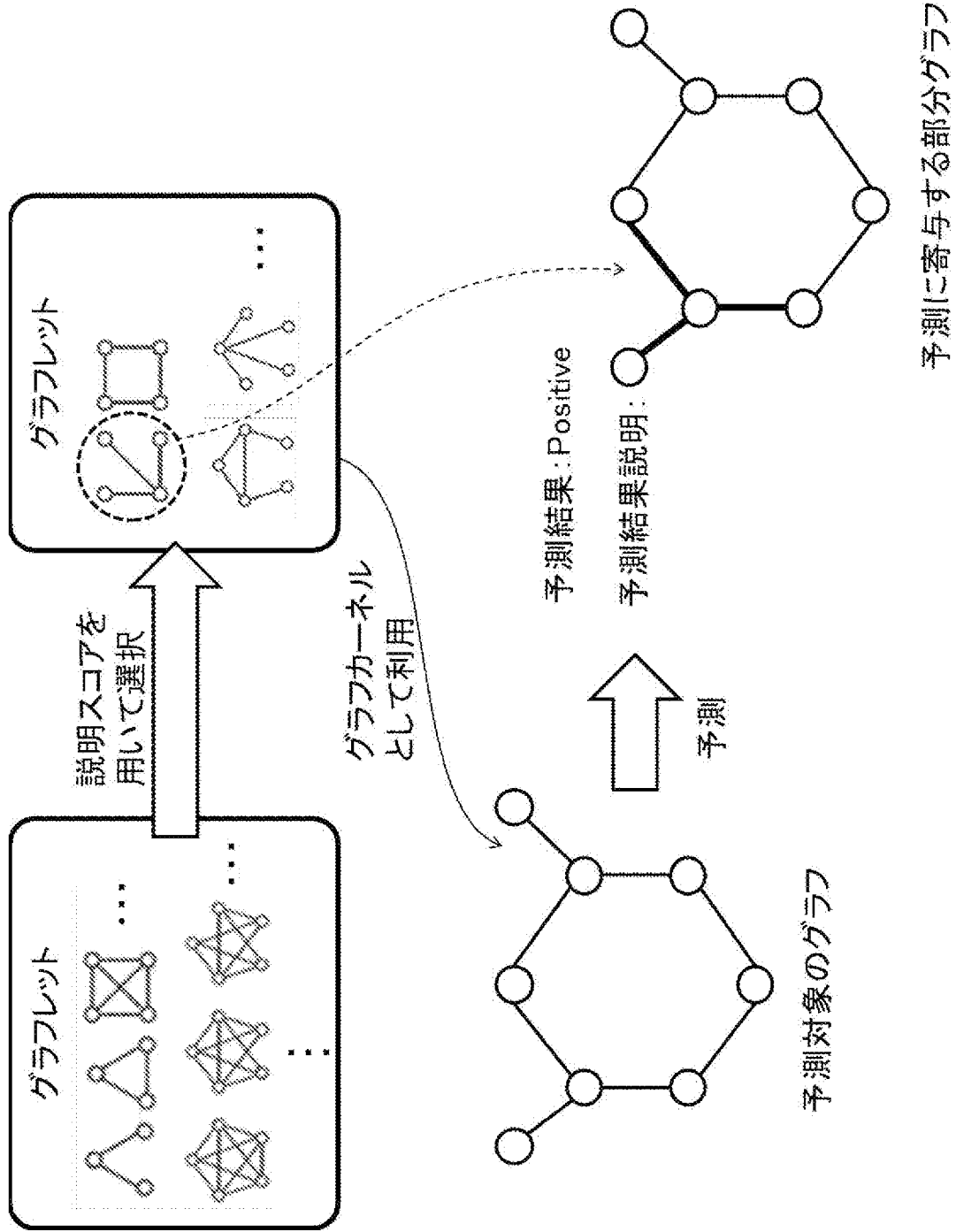
[図6]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/016245

A. CLASSIFICATION OF SUBJECT MATTER		
G06N 20/00(2019.01)i FI: G06N20/00		
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) G06N3/00-99/00		
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 1971-2022 Registered utility model specifications of Japan 1996-2022 Published registered utility model applications of Japan 1994-2022		
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	WU, Jia et al. Boosting for Multi-Graph Classification. IEEE Transactions on Cybernetics [online]. 08 July 2014 [retrieved on 14 June 2022], vol. 45, Issue 3, pp. 430-443, <URL: https://ieeexplore.ieee.org/abstract/document/6848779 > entire text	1-16
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> 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 14 June 2022		Date of mailing of the international search report 21 June 2022
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00		
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）） G06N3/00-99/00 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 1971-2022年 日本国実用新案登録公報 1996-2022年 日本国登録実用新案公報 1994-2022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	WU, Jia, et al., Boosting for Multi-Graph Classification, IEEE Transactions on Cybernetics [online], 2014.07.08 [検索日: 2022.06.14], Volume: 45, Issue: 3, pp.430-443, <URL: https://ieeexplore.ieee.org/abstract/document/6848779> 全文	1-16
.....		
<input type="checkbox"/> C欄の続きにも文献が列挙されている。		
<input type="checkbox"/> 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 14.06.2022	国際調査報告の発送日 21.06.2022	
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） 杉浦 孝光 5B 5287 電話番号 03-3581-1101 内線 3545	