

(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
2022年9月15日(15.09.2022)



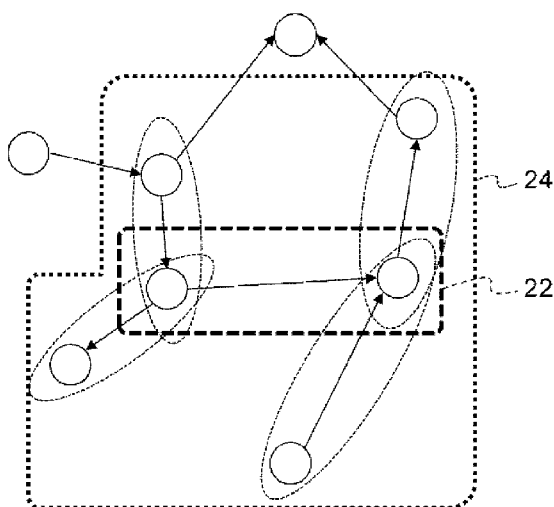
(10) 国際公開番号

WO 2022/190215 A1

- (51) 国際特許分類:
G06N 20/00 (2019.01)
- (21) 国際出願番号: PCT/JP2021/009344
- (22) 国際出願日: 2021年3月9日(09.03.2021)
- (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: 鵜飼 孝典 (UKAI, Takanori); 〒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 HK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, 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: MACHINE LEARNING PROGRAM, PREDICTION PROGRAM, DEVICE, AND METHOD

(54) 発明の名称: 機械学習プログラム、予測プログラム、装置、及び方法



(57) Abstract: This machine learning device adds additional data indicated by a triple (22) that includes a first node, a second node, and a first edge indicating a relationship between the first node and the second node, to graph data which includes a plurality of nodes and a plurality of edges indicating relationships between the plurality of nodes and in which a plurality of embedding vectors respectively indicating the plurality of nodes and the plurality of edges are calculated by machine learning, so as to update the embedding vectors. In doing so, the machine learning device updates, in the graph data, the embedding vectors respectively for one or more nodes and one or more edges which are in connections to the triple (22) indicating the additional data and which are included in a range (24) where the connections are made in a specific condition.

(57) 要約: 機械学習装置は、第1のノードと第2のノードと第1のノードと第2のノードとの関係を示す第1のエッジとを含むトリプル(22)で表される追加データを、複数のノードと複数のノード間の関係を示す複数のエッジとを含むグラフデータであって、機械学習によって複数のノードと複数のエッジとのそれぞれを表す複数の埋め込みベクトルが計算されているグラフデータに追加して埋め込みベクトルを更新する際、グラフデータにおいて、追加データを表すトリプル(22)と特定の条件で接続される範囲(24)の一又は複数のノードと一又は複数のエッジとのそれぞれの埋め込みベクトルを更新する。



WO 2022/190215 A1

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))

明 細 書

発明の名称：

機械学習プログラム、予測プログラム、装置、及び方法

技術分野

[0001] 開示の技術は、機械学習技術に関する。

背景技術

[0002] 従来、複数のノードと、ノード間の関係を示す複数のエッジとを含むグラフデータを用いて、特定の事象等の予測を行う技術が提案されている。例えば、予測される部位特異的なタンパク質リン酸化陳述の発見のためのシステムが提案されている。このシステムは、既知の部位特異的な陳述を記憶しているリン酸化データセット、キナーゼ・ファミリー・データベース、及びタンパク質配列データベースからのデータを使って、モチーフ・ベースのリン酸化陳述を生成する。また、このシステムは、モチーフ・ベースの陳述から陰性陳述を生成し、モチーフ・ベースの陳述を組み合わせることによって候補陳述を生成し、候補陳述の陰性陳述との組み合わせによって陰性候補陳述を生成する。また、このシステムは、候補陳述及び陰性候補陳述を統計的關係学習モデルにおいて使って候補陳述にスコア付けし、予測されるモチーフ・ベースのリン酸化陳述を保持する。また、このシステムは、タンパク質配列データベースからのデータ入力および一般化されたフォーマットでの予測されたモチーフ・ベースのリン酸化陳述を受け取り、部位特異的なフォーマットで予測された部位特異的なリン酸化陳述を生成する。

[0003] また、例えば、それぞれ一对のエンティティの間の関係についての情報を与える複数の述語、及び関連するエンティティから知識グラフを完成させる装置が提案されている。この装置は、複数の述語及び関連するエンティティを含む入力を受領し、公準データベースを検索して、複数の述語のうちで互いに等価である又は互いの逆である述語を識別する。また、この装置は、公準データベース及び識別された述語を使って、互いに関係しているさらなる

述語を識別し、識別された述語及び関連するエンティティをベクトル空間に埋め込んで知識グラフを完成させる。

[0004] また、例えば、知識グラフに基づいてユーザのためのレコメンデーションを生成するシステムが提案されている。このシステムは、複数のエンティティと複数のエンティティの属性を含むデータセットを収集し、複数のエンティティ間の関係性を生成し、複数のエンティティ、属性、及び関係性を知識グラフに格納する。また、このシステムは、知識グラフに格納される複数のエンティティ、属性、及び関係性の表現を学習し、新たなエンティティ及び新たなエンティティの属性のゼロショット学習を実行し、新たなエンティティ及び新たなエンティティの属性を知識グラフに格納する。そして、このシステムは、知識グラフに基づいてユーザのためのレコメンデーションを生成する。

先行技術文献

特許文献

- [0005] 特許文献1：特開2018-206374号公報
特許文献2：特開2018-85116号公報
特許文献3：特開2019-125364号公報

発明の概要

発明が解決しようとする課題

[0006] グラフデータを構成するエンティティ及びエンティティ間の関係を表すリレーションのトリプルに基づいた機械学習により、エンティティ及びリレーションの各々を表す埋め込みベクトルが生成されているとする。このグラフデータに対して、新たに得られたトリプルの情報を反映するために、新たに得られたトリプルを追加したグラフデータに対して再度一から機械学習を実行する場合、すなわちバッチ処理を実行する場合、機械学習の処理負荷が高いという問題がある。

[0007] 一つの側面として、開示の技術は、グラフデータの埋め込みベクトルの機

械学習の処理負荷を低減することを目的とする。

課題を解決するための手段

[0008] 一つの態様として、開示の技術は、第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報を取得する。また、開示の技術は、前記情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、ベクトルを更新する。更新するベクトルは、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルである。

発明の効果

[0009] 一つの側面として、グラフデータの埋め込みベクトルの機械学習の処理負荷を低減することができる、という効果を有する。

図面の簡単な説明

- [0010] [図1]事例データの一例を示す図である。
[図2]グラフデータの一例を示す図である。
[図3]埋め込みベクトルの計算を説明するための図である。
[図4]追加データを説明するための図である。
[図5]機械学習装置の機能ブロック図である。
[図6]追加データを表すトリプルを説明するための図である。
[図7]機械学習済みのグラフデータへの追加データの追加を説明するための図である。
[図8]第1実施形態におけるオンライン学習の範囲の決定を説明するための図である。
[図9]機械学習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10]機械学習処理の一例を示すフローチャートである。
[図11]第1実施形態における範囲決定処理の一例を示すフローチャートであ

る。

[図12]予測処理の一例を示すフローチャートである。

[図13]第2実施形態におけるオンライン学習の範囲の決定を説明するための図である。

[図14]第2実施形態におけるオンライン学習の範囲の決定を説明するための図である。

[図15]第2実施形態における範囲決定処理の一例を示すフローチャートである。

[図16]第2実施形態におけるオンライン学習の範囲の決定を説明するための図である。

[図17]予測対象となるエッジ及び追加データについて説明するための図である。

[図18]第3実施形態におけるオンライン学習の範囲の決定を説明するための図である。

[図19]第3実施形態におけるオンライン学習の範囲の決定を説明するための図である。

発明を実施するための形態

[0011] <グラフデータにおける埋め込みベクトルの機械学習について>

以下の各実施形態に係る機械学習装置は、埋め込みベクトルを機械学習済みのグラフデータに対し、新たにデータが取得された場合に、オンライン学習によりグラフデータを更新する。以下では、埋め込みベクトルを機械学習済みのグラフデータを「機械学習済みグラフデータ」ともいう。また、新たに取得されるデータを「追加データ」という。各実施形態の詳細を説明する前に、機械学習済みのグラフデータ、及びグラフデータのオンライン学習として一般的に想定される手法の問題点について説明する。

[0012] まず、医薬品の投与における、想定以外の作用（以下、「副作用」という）の予測に用いられるグラフデータを例に、機械学習済みのグラフデータについて説明する。ここでは、情報処理装置が機械学習済みのグラフデータを

生成するものとして説明する。情報処理装置は、以下で説明する各実施形態に係る機械学習装置であってもよいし、機械学習装置以外の他のコンピュータであってもよい。

[0013] 情報処理装置は、事例データを、ノードとノード間を接続するエッジとで構成されるグラフに変換したグラフデータを生成する。ここでの例の場合、事例データは、患者の属性、投与した医薬品、患者が罹患している疾病等の情報、及び、副作用の情報を含むデータである。図1に、事例データの一例を示す。図1の例では、患者毎に、「ID」、「性別」、「年代」、「体重」、「身長」、「医薬品」、「疾病」、及び「副作用」の情報が含まれている。「ID」は、患者の識別情報である。「性別」、「年代」、「体重」、及び「身長」は、患者の属性の一例である。「医薬品」は、患者に投与された医薬品の名称である。「疾病」は、患者が罹患している原疾病の名称である。「副作用」は、「医薬品」に示す医薬品を投与した際に発生した副作用の情報である。

[0014] 情報処理装置は、上記のような事例データに含まれる各項目の各値を示すノードを生成し、各「ID」のノードから、そのIDが示す患者についての属性、医薬品、疾病、及び副作用の各々を示すノードへエッジを接続することで、グラフデータを生成する。図2に、図1に示す事例データから生成されたグラフデータの一例を示す。図2において、中に各値が表記された丸で示すノードは、属性、医薬品、及び疾病の各々を示すノードであり、中に副作用が表記された角丸の四角で示すノードは、副作用を示すノードであり、ノード間を接続する矢印がエッジである。

[0015] また、図2の例では、事例データを示すグラフデータに、背景知識を示すオントロジー（図2中の破線部）を接続した場合を示している。背景知識を示すオントロジーとは、対象の分野における背景知識を体系化したものであり、ここでは、例えば、疾病の類似性や関係、医薬品の類似性や含まれている成分等の情報を木構造等の形式で整理したものである。図2の例では、オントロジーに関するノードを楕円で表している。事例データをグラフデータ

に変換することで、このようなオントロジーの接続が容易となる。疾病が類似する場合や、同じ成分が含まれている医薬品を投与する場合などには、同じような副作用が起こる可能性があるため、事例データにオントロジーを接続することで、副作用の予測の精度が向上する。

[0016] 情報処理装置は、生成したグラフデータに含まれる各ノード及び各エッジを表現する埋め込みベクトルを計算する。具体的には、情報処理装置は、グラフデータに含まれるノード及びエッジの各々を n 次元のベクトル空間に写像することにより埋め込みベクトルを計算する。より具体的に、図3上段の図に示すように、ノードA、B、C、ノードA-B間のエッジ r 、及びノードC-B間のエッジ r を含むグラフデータを例に、埋め込みベクトルの計算について説明する。ここでは、説明を簡単にするため、2次元のベクトル空間に写像する場合について説明する。

[0017] まず、情報処理装置は、図3中段の図に示すように、グラフデータに含まれるノード及びエッジの各々を、初期値のベクトルとしてベクトル空間に配置する。そして、情報処理装置は、ノードの接続関係を表現するように、各ベクトルの配置を最適化する。図3の例では、情報処理装置は、図3下段の図に示すように、ベクトルA + ベクトル r がベクトルBに近くなるように、かつ、ベクトルC + ベクトル r がベクトルBに近くなるように、各ベクトルの配置を最適化する。最適化後のベクトルが、そのベクトルが示すノードの埋め込みベクトルとなる。情報処理装置により、各ノード及び各エッジについての埋め込みベクトルを計算されたグラフデータが、機械学習済みのグラフデータとなる。

[0018] 上記のようなグラフデータに対して、追加データが取得された場合、機械学習済みのグラフデータを更新し、追加データが表す変化に追従する必要がある。グラフデータの更新の手法としては、上記と同様に、追加データを追加したグラフデータの各ノード及び各エッジの埋め込みベクトルを全て計算し直す手法、すなわちバッチ処理により機械学習を実行する手法が考えられる。しかし、バッチ処理による機械学習は処理負荷が高い。特に、グラフデ

ータに含まれるノード数及びエッジ数が膨大な場合や、追加データが取得される頻度が高い場合には、より効率的にグラフデータを更新する手法が望ましい。

[0019] そこで、追加データが取得される都度、追加データのみを使って機械学習を実行し、グラフを更新するオンライン学習の手法が考えられる。すなわち、情報処理装置が、埋め込みベクトルの計算を1から計算し直してグラフデータを更新するのではなく、追加データによる機械学習を実行して、機械学習済みのグラフデータのパラメータである埋め込みベクトルを随時更新していく。この場合、1回の機械学習当たりの処理負荷が低く、過去の機械学習の実行で使用したデータを全て蓄えておく必要がなく、また、データの変化にすぐに対応できるという利点がある。

[0020] 通常のオンライン学習では、追加データは全ての素性を持っているため、機械学習済みのモデルの全てのパラメータを更新することができる。しかし、グラフデータの埋め込みベクトルのオンライン学習では、追加データによって変化するのは、その追加データが表すノード及びエッジの部分のみである。例えば、図2に示す機械学習済みのグラフデータに対して、図4に破線部に示すように、ノード「男性」とノード「メトホルミン塩酸塩」との間にエッジが追加されたとする。この場合、この2つのノードとエッジとからなるトリプル（3つ組）に関する埋め込みベクトルのみが更新される。このように、オンライン学習により更新される範囲が、追加データの範囲に限定される場合には、グラフデータを用いた予測に影響を与えるだけの十分な変化が得られない。そこで、以下の各実施形態では、オンライン学習を実行する範囲を追加データに基づいて決定した上で、オンライン学習を実行する。これにより、各実施形態は、オンライン学習の利点を生かしつつ、バッチ処理による機械学習の場合よりも処理負荷を低減する。以下、開示の技術に係る実施形態の一例を説明する。

[0021] <第1実施形態>

図5に示すように、機械学習装置10は、機能的には、取得部12と、更

新部 14 と、予測部 18 とを含む。また、機械学習装置 10 の所定の記憶領域には、グラフデータ 16 が記憶されている。グラフデータ 16 は、追加データが入力され、更新処理が実行される前の状態の機械学習済みのグラフデータである。なお、更新処理が実行されたグラフデータをグラフデータ 20 とする。

[0022] 取得部 12 は、機械学習装置 10 に入力された追加データを取得する。追加データは、図 6 に示すように、RDF (Resource Description Framework) 形式のデータであり、2つのエンティティと、エンティティ間のリレーションとの3つの要素を含むトリプルの情報である。具体的には、図 6 に示すように、追加データは、サブジェクトとなるノード、オブジェクトとなるノード、及びサブジェクトとオブジェクトとのリレーションとなるエッジで構成される。エッジの始点側のノードがサブジェクトとなるノードであり、エッジの終点側のノードがオブジェクトとなるノードである。サブジェクトとなるノード及びオブジェクトとなるノードは、開示の技術の「第1のノード」及び「第2のノード」の一例であり、リレーションとなるエッジは、開示の技術の「第1のエッジ」の一例である。また、追加データを表すトリプルは、開示の技術の「第1のトリプル」の一例である。取得部 12 は、取得した追加データを更新部 14 へ受け渡す。

[0023] 更新部 14 は、追加データと、機械学習済みのグラフデータ 16 の複数のノードと複数のエッジとのそれぞれを表す複数の埋め込みベクトルとに基づいて、複数の埋め込みベクトルのうち、特定の埋め込みベクトルを更新する。特定の埋め込みベクトルは、オンライン学習の範囲として決定された範囲に含まれるノード及びエッジを表す埋め込みベクトルである。具体的には、特定の埋め込みベクトルは、機械学習済みのグラフデータ 16 において、追加データを表すトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれの埋め込みベクトルである。更新部 14 は、例えば、機械学習済みのグラフデータ 16 に追加データを表すトリプルを追加した場合に、追加データを表すトリプルと接続されるトリプルであって、

追加データを表すトリプルとの距離が所定値以下であることを、特定の条件としてよい。ここでの距離は、ノード間のエッジ数である。すなわち、あるノードから他のノードまでの距離は、2つのノードをつなぐパスに含まれるエッジ数である。

[0024] 例えば、図7上図に示すような機械学習済みのグラフデータ16に、図7下図の破線部に示すような追加データを表すトリプル22が追加されたとする。上記の特定の条件の所定値が、エッジ数1で表される距離であるとする。この場合、更新部14は、図8に示すように、グラフデータ16に含まれるトリプルのうち、追加データを表すトリプル22と、追加データを表すトリプル22に含まれるノードを要素とするトリプルとを、オンライン学習の範囲24として決定する。図8において、点線の楕円で囲ったトリプルが、追加データを表すトリプル22に含まれるノードを要素とするトリプルである。なお、上記の所定値は、エッジ数1で表される距離に限定されず、グラフデータ16の規模や種類等に応じて適宜設定すればよい。

[0025] 更新部14は、決定したオンライン学習の範囲24に含まれるトリプルの各々について、図3を参照して説明したように、トリプルに含まれるノード及びエッジの各々の埋め込みベクトルを計算する。これにより、機械学習済みのグラフデータ16のうち、範囲24に含まれるノード及びエッジの埋め込みベクトルが更新され、更新されたグラフデータ20となる。

[0026] 予測部18は、更新されたグラフデータ20を用いて、入力データにおける、予測対象となるエッジの有無を予測する。入力データは、図1に示すような事例データのうち、予測対象の項目（図1の例では「副作用」）を含まないデータである。例えば、予測部18は、入力データをグラフデータに変換し、上記と同様に、入力データを表すグラフデータに含まれるノード及びエッジの各々の埋め込みベクトルを計算する。そして、予測部18は、グラフデータ20の埋め込みベクトルと、入力データの埋め込みベクトルとの類似性に基づいて、グラフデータ20において、予測対象となるエッジに接続する可能性のあるノードを特定する。そして、予測部18は、特定したノード

ドに予測対象となるエッジが接続されているか否かを予測結果として出力する。例えば、予測部18は、特定したノードに予測対象となるエッジが接続されている場合は、TRUEを示す予測結果、接続されていない場合は、FALSEを示す予測結果を出力する。

[0027] 図4の例の場合、副作用を示すノード「中毒性皮疹」に接続するエッジが予測対象のエッジであり、「ID」を示すノードの各々が、予測対象となるエッジに接続する可能性のあるノードである。そして、予測部18は、特定した患者の「ID」を示すノードと、副作用を示すノード「中毒性皮疹」とを接続するエッジの有無により、予測対象の患者が副作用「中毒性皮疹」を発症する可能性があるか否かを予測する。

[0028] なお、グラフデータを用いた予測の特徴として、入力データは、事例データに含まれる、予測対象の項目以外の全ての項目の値を有している必要はない。一部の項目の値が欠ける入力データであってもグラフデータを生成可能であり、生成された入力データのグラフデータと、グラフデータ20の一部との類似性を判断できれば、予測が可能のためである。

[0029] 機械学習装置10は、例えば図9に示すコンピュータ40で実現することができる。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。CPU41、メモリ42、記憶部43、入出力装置44、R/W部45、及び通信I/F46は、バス47を介して互いに接続される。

[0030] 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できる。記憶媒体としての記憶部43には、コンピュータ40を、機械学習装置10として機能させるための機械学習プログラム50が記憶される。機械学習プログラム50は、取得プロセ

ス52と、更新プロセス54と、予測プロセス58とを有する。また、記憶部43は、グラフデータ16(20)を構成する情報が記憶される情報記憶領域60を有する。

[0031] CPU41は、機械学習プログラム50を記憶部43から読み出してメモリ42に展開し、機械学習プログラム50が有するプロセスを順次実行する。CPU41は、取得プロセス52を実行することで、図5に示す取得部12として動作する。また、CPU41は、更新プロセス54を実行することで、図5に示す更新部14として動作する。また、CPU41は、予測プロセス58を実行することで、図5に示す予測部18として動作する。また、CPU41は、情報記憶領域60から情報を読み出して、グラフデータ16(20)をメモリ42に展開する。これにより、機械学習プログラム50を実行したコンピュータ40が、機械学習装置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CPU41はハードウェアである。また、CPU41は、開示の技術の「制御部」の一例である。

[0032] なお、機械学習プログラム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。

[0033] 次に、第1実施形態に係る機械学習装置10の作用について説明する。機械学習装置10に追加データが入力されると、機械学習装置10において、図10に示す機械学習処理が実行される。機械学習装置10に入力データが入力されると、機械学習装置10において、図12に示す予測処理が実行される。なお、機械学習処理は、開示の技術の機械学習方法の一例であり、予測処理は、開示の技術の予測方法の一例である。以下、機械学習処理及び予測処理について詳述する。

[0034] まず、図10に示す機械学習処理について説明する。ステップS10で、取得部12は、機械学習装置10に入力された追加データを取得し、取得した追加データを更新部14へ受け渡す。そして、更新部14が、追加データを表すトリプル(h__add, r__add, t__add)の各埋め込みベク

トル V_{h_add} 、 V_{r_add} 、 V_{t_add} を特定する。具体的には、更新部 14 は、トリプル (h_add , r_add , t_add) をグラフデータ 16 に追加した場合に、サブジェクトであるノード h_add に対応するグラフデータ 16 のノードについての埋め込みベクトルを V_{h_add} として特定する。同様に、更新部 14 は、オブジェクトであるノード t_add に対応するグラフデータ 16 のノードについての埋め込みベクトルを V_{t_add} として特定する。また、更新部 14 は、エッジ r_add と同様のリレーションを示すグラフデータ 16 のエッジについての埋め込みベクトルを V_{r_add} として特定する。

[0035] 次に、ステップ S 12 で、更新部 14 が、 $V_{h_add} + V_{r_add} - V_{t_add}$ が所定の閾値 TH より小さいか否かを判定する。閾値 TH としては、 $V_{h_add} + V_{r_add}$ と V_{t_add} とが概ね一致していると判定可能な値を予め設定しておく。 $V_{h_add} + V_{r_add} - V_{t_add} < TH$ の場合、追加データを追加しても、グラフデータ 16 の埋め込みベクトルは変化しないため、そのまま機械学習処理を終了する。一方、 $V_{h_add} + V_{r_add} - V_{t_add} \geq TH$ の場合、グラフデータ 16 の埋め込みベクトルを追加データに応じて更新する必要があるため、ステップ S 20 へ移行する。

[0036] ステップ S 20 では、範囲決定処理が実行される。ここで、図 11 を参照して、範囲決定処理について説明する。ここでは、追加データを表すトリプルと直接接続されるトリプル（距離がエッジ数 1 のトリプル）がオンライン学習の範囲に含まれるように範囲を決定する場合について説明する。ステップ S 22 で、更新部 14 が、オンライン学習の範囲に含まれるグラフデータ 16 のトリプルを格納するための集合 T として空集合を作成する。次に、ステップ S 24 で、更新部 14 が、追加データを表すトリプル (h_add , r_add , t_add) を集合 T に追加する。次に、ステップ S 26 で、更新部 14 が、グラフデータ 16 から、追加データを表すトリプル 22 と直接接続されるトリプル、すなわち、追加データを表すトリプル 22 に含まれるノードを要素とするトリプルを全て取り出し、集合 T に追加する。集合 T に追加されたトリプルを、トリプル (h_add , r_add , t_add)

)も含め、以下では (h, r, t) と表記する。更新部14は、集合Tに追加されたトリプルを構成するノード及びエッジの各々にランダムに埋め込みベクトルの初期値を与える。そして、機械学習処理(図10)に戻る。

[0037] 次に、ステップS30で、後述するステップS32~S40の更新処理を所定回数(N回)繰り返したか否かを判定する。繰り返し回数がN回に到達していない場合には、ステップS32へ移行する。ステップS32では、更新部14が、集合Tから1つのトリプル (h, r, t) を取り出す。次に、ステップS34で、更新部14が、取り出したトリプル (h, r, t) の各要素の埋め込みベクトル V_h 、 V_r 、及び V_t を用いて、 $V_h + V_r - V_t < TH$ か否かを判定する。 $V_h + V_r - V_t < TH$ の場合、ステップS30に戻り、 $V_h + V_r - V_t \geq TH$ の場合、ステップS36へ移行する。

[0038] ステップS36では、更新部14が、 $V_h + V_r - V_t < 0$ か否かを判定する。 $V_h + V_r - V_t < 0$ の場合、ステップS38へ移行し、更新部14が、 $(V_h + V_r - V_t)$ に所定の係数 α を乗算した値を V_h に加算することにより、ノードhについての埋め込みベクトルを更新する。一方、 $V_h + V_r - V_t \geq 0$ の場合、ステップS40へ移行し、更新部14が、 $(V_h + V_r - V_t)$ に所定の係数 α を乗算した値を V_h から減算することにより、ノードhについての埋め込みベクトルを更新する。そして、処理はステップS30に戻る。ステップS38及びS40をN回繰り返す処理は、図3を参照して説明した埋め込みベクトルの計算における最適化に相当する。

[0039] ステップS30で、繰り返し回数がN回に到達したと判定されると、機械学習処理は終了する。これにより、上記ステップS20の範囲決定処理で決定された範囲のノード及びエッジの埋め込みベクトルが更新されたグラフデータ20が生成される。

[0040] 次に、図12に示す予測処理について説明する。ステップS50で、予測部18が、機械学習装置10に入力された入力データを取得し、グラフデータに変換し、入力データを示すグラフデータに含まれるノード及びエッジの各々の埋め込みベクトルを計算する。次に、ステップS52で、予測部18

が、グラフデータ20の埋め込みベクトルと、入力データの埋め込みベクトルとの類似性に基づいて、グラフデータ20において、予測対象となるエッジに接続する可能性のあるノードを特定する。次に、ステップS54で、予測部18が、特定したノードに予測対象となるエッジが接続されているか否かを予測結果として出力し、予測処理は終了する。

[0041] 以上説明したように、第1実施形態に係る機械学習装置は、複数のノードと複数のノード間の関係を示す複数のエッジとを含み、機械学習によってノード及びエッジのそれぞれを表す埋め込みベクトルが計算されたグラフデータに、追加データを追加する。そして、機械学習装置は、グラフデータの埋め込みベクトルを更新する。この際、機械学習装置は、グラフデータにおいて、追加データを表すトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとを含む範囲を、オンライン学習の範囲として決定する。そして、機械学習装置は、決定した範囲のノード及びエッジのそれぞれを表す埋め込みベクトルを更新する。これにより、グラフデータの埋め込みベクトルの機械学習の処理負荷を低減することができる。

[0042] <第2実施形態>

次に、第2実施形態について説明する。なお、第2実施形態に係る機械学習装置において、第1実施形態に係る機械学習装置10と同様の構成については、同一符号を付して詳細な説明を省略する。また、第1実施形態と第2実施形態とで、符号の末尾2桁が共通する機能構成において、共通する機能についての詳細な説明を省略する。

[0043] 図5に示すように、機械学習装置210は、機能的には、取得部12と、更新部214と、予測部18とを含む。また、機械学習装置210の所定の記憶領域には、グラフデータ16(20)が記憶される。

[0044] 更新部214は、第1実施形態における更新部14と同様に、グラフデータ16の埋め込みベクトルを追加データに基づいて更新し、更新されたグラフデータ20を生成する。第2実施形態における更新部214は、オンライン学習の範囲の決定方法が、第1実施形態における更新部14とは異なるた

め、以下、第2実施形態における更新部214による範囲の決定方法について説明する。

[0045] 更新部214は、グラフデータ16の構造に応じて、オンライン学習の範囲を決定する。例えば、図1に示すようなテーブル形式のデータを変換したグラフデータと、SNS (Social Networking Service) におけるユーザのつながりを表すようなグラフデータとでは、グラフデータ16の構造が異なる。前者のグラフデータ16の場合、例えば図2に示すように、「ID」を示す1つのノードから複数のノードへエッジが接続される傾向が強い。一方、後者のグラフデータ16の場合、1つのノードから連鎖的に他のノードにエッジが接続される傾向が強い。このような構造の違いにより、グラフデータ16に追加データを追加した際の影響範囲も異なる。例えば、前者のグラフデータ16の場合、追加データの影響が、ノードの兄弟関係（同一の親ノードに接続する子ノード同士）の方向に強く働き、後者のグラフデータ16の場合、子孫関係の方向に強く働くと考えられる。

[0046] そこで、更新部214は、上記のようなグラフデータの構造を判定する。具体的には、更新部214は、グラフデータ16のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合を算出する。この割合が少ない場合には、上記の前者のグラフデータの構造を表し、この割合が多い場合には、上記の後者のグラフデータの構造を表す。

[0047] 更新部214は、算出した割合が所定割合（例えば、50%）以下の場合、図13に示すように、追加データを表すトリプル22においてサブジェクトであるノードをオブジェクトとして接続するトリプル24Aが範囲24に含まれるように決定する。また、更新部214は、追加データを表すトリプル22においてオブジェクトであるノードをサブジェクトとして接続するトリプル24Bが範囲24に含まれるように決定する。すなわち、更新部214は、追加データを表すトリプル22と、トリプル22に接続するトリプル24A、24Bとに含まれるノード間の関係が子孫関係となるように、範囲24を決定する。一方、更新部214は、算出した割合が所定割合（例えば

、50%)を超える場合、図14に示すように、追加データを表すトリプル22においてサブジェクトであるノードをサブジェクトとして接続するトリプル24Cが範囲24に含まれるように決定する。また、更新部214は、追加データを表すトリプル22においてオブジェクトであるノードをオブジェクトとして接続するトリプル24Dが範囲24に含まれるように決定する。すなわち、更新部214は、追加データを表すトリプル22と、トリプル22に接続するトリプル24C、24Dとに含まれるノード間の関係が、共通するノードを中心として、兄弟関係となるように、範囲24を決定する。

[0048] 機械学習装置210は、例えば図9に示すコンピュータ40で実現することができる。コンピュータ40の記憶部43には、コンピュータ40を、機械学習装置210として機能させるための機械学習プログラム250が記憶される。機械学習プログラム250は、取得プロセス52と、更新プロセス254と、予測プロセス58とを有する。また、記憶部43は、グラフデータ16(20)を構成する情報が記憶される情報記憶領域60を有する。

[0049] CPU41は、機械学習プログラム250を記憶部43から読み出してメモリ42に展開し、機械学習プログラム250が有するプロセスを順次実行する。CPU41は、更新プロセス254を実行することで、図5に示す更新部214として動作する。他のプロセスについては、第1実施形態に係る機械学習プログラム50と同様である。これにより、機械学習プログラム250を実行したコンピュータ40が、機械学習装置210として機能することになる。なお、機械学習プログラム250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等で実現することも可能である。

[0050] 次に、第2実施形態に係る機械学習装置210の作用について説明する。第2実施形態では、機械学習処理における範囲決定処理が第1実施形態と異なるため、図15を参照して、第2実施形態における範囲決定処理について説明する。なお、第2実施形態における範囲決定処理において、第1実施形態における範囲決定処理(図11)と同様の処理については、同一のステップ番号を付して詳細な説明を省略する。

- [0051] ステップS 2 2 2 及びS 2 4 を経て、ステップS 2 2 0 へ移行すると、更新部2 1 4 が、グラフデータ1 6 のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合を算出する。次に、ステップS 2 2 2 で、更新部2 1 4 が、算出した割合が所定割合（例えば、5 0 %）以下か否かを判定する。算出した割合が所定割合以下の場合、ステップS 2 2 4 へ移行し、所定割合を超える場合、ステップS 2 2 8 へ移行する。
- [0052] ステップS 2 2 4 では、更新部2 1 4 が、グラフデータ1 6 から、オブジェクトがノードh__a d d（サブジェクトであるノード）のトリプル（h, r, t）を全て取り出して、集合Tに追加する。次に、ステップS 2 2 6 で、更新部2 1 4 が、グラフデータ1 6 から、サブジェクトがノードt__a d d（オブジェクトであるノード）のトリプル（h, r, t）を全て取り出して、集合Tに追加し、ステップS 2 3 2 へ移行する。
- [0053] 一方、ステップS 2 2 8 では、更新部2 1 4 が、グラフデータ1 6 から、サブジェクトがノードh__a d d（サブジェクトであるノード）のトリプル（h, r, t）を全て取り出して、集合Tに追加する。次に、ステップS 2 3 0 で、更新部2 1 4 が、グラフデータ1 6 から、オブジェクトがノードt__a d d（オブジェクトであるノード）のトリプル（h, r, t）を全て取り出して、集合Tに追加し、ステップS 2 3 2 へ移行する。
- [0054] ステップS 2 3 2 では、更新部2 1 4 が、取り出したトリプルのノードのうち、ノードh__a d d 及びt__a d d 以外のノード間で、リレーションがr__a d d と同一のトリプル（h, r, t）を全て取り出して、集合Tに追加する。これは、図1 6 に示すトリプル2 4 E のようなトリプルを範囲2 4 に含めるものである。そして、範囲決定処理は終了し、機械学習処理（図1 0）にリターンする。
- [0055] 以上説明したように、第2 実施形態に係る機械学習装置は、グラフデータのノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合に応じて、オンライン学習の範囲を決定する。これにより、グラフデータの構造に応じて、より適切にオンライン学習の範囲を決定でき、グラフデ

ータの埋め込みベクトルの機械学習の処理負荷を低減することができる。

[0056] <第3実施形態>

次に、第3実施形態について説明する。なお、第3実施形態に係る機械学習装置において、第1実施形態に係る機械学習装置10と同様の構成については、同一符号を付して詳細な説明を省略する。また、第1実施形態と第3実施形態とで、符号の末尾2桁が共通する機能構成において、共通する機能についての詳細な説明を省略する。

[0057] 図5に示すように、機械学習装置310は、機能的には、取得部12と、更新部314と、予測部18とを含む。また、機械学習装置310の所定の記憶領域には、グラフデータ16(20)が記憶される。

[0058] 更新部314は、第1実施形態における更新部14と同様に、グラフデータ16の埋め込みベクトルを追加データに基づいて更新し、更新されたグラフデータ20を生成する。第3実施形態における更新部314は、オンライン学習の範囲の決定方法が、第1実施形態における更新部14とは異なるため、以下、第3実施形態における更新部314による範囲の決定方法について説明する。

[0059] ここで、図17に示すように、グラフデータを用いた予測において、予測対象となるエッジ26の有無を予測する場合、追加データにより予測に影響が及ぶようにオンライン学習の範囲を決定することが望ましい。しかし、例えば、追加データを表すトリプルを追加した部分と、予測対象となるエッジ26までの距離が離れている場合等、第1及び第2実施形態におけるオンライン学習の範囲の決定方法では、予測に影響が及ぶように範囲が決定されない場合がある。

[0060] そこで、更新部314は、追加データを表すトリプル22のノードのうち、予測対象となるエッジ26に接続する可能性のあるノード(以下「予測ノード」という)28までの距離が短い一方のノードから予測ノード28までのパスが含まれるように範囲を決定する。例えば、更新部314は、図13に示すように範囲24を決定した場合、トリプル22のサブジェクトとなる

ノードから予測ノード28までの最短パス、及びオブジェクトとなるノードから予測ノード28までの最短パスを探索し、距離が短い方のパスを選択する。この場合、図18に示すように、更新部314は、サブジェクトとなるノードから予測ノード28までの最短パスを選択する。更新部314は、図18に示すように、選択したパスに含まれるノード及びエッジが含まれるように、範囲24を拡張する。

[0061] 機械学習装置310は、例えば図9に示すコンピュータ40で実現することができる。コンピュータ40の記憶部43には、コンピュータ40を、機械学習装置310として機能させるための機械学習プログラム350が記憶される。機械学習プログラム350は、取得プロセス52と、更新プロセス354と、予測プロセス58とを有する。また、記憶部43は、グラフデータ16(20)を構成する情報が記憶される情報記憶領域60を有する。

[0062] CPU41は、機械学習プログラム350を記憶部43から読み出してメモリ42に展開し、機械学習プログラム350が有するプロセスを順次実行する。CPU41は、更新プロセス354を実行することで、図5に示す更新部314として動作する。他のプロセスについては、第1実施形態に係る機械学習プログラム50と同様である。これにより、機械学習プログラム350を実行したコンピュータ40が、機械学習装置310として機能することになる。なお、機械学習プログラム350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等で実現することも可能である。

[0063] 第3実施形態に係る機械学習装置310の作用については、機械学習処理における範囲決定処理が、第1及び第2実施形態における範囲決定処理(図11、図15)と異なるだけであるため、詳細な説明を省略する。具体的には、更新部314が、第1及び第2実施形態における範囲決定処理で決定される範囲に、追加データを表すトリプル22から予測ノード28までのノード及びエッジが含まれるように範囲を決定する点が異なる。

[0064] 以上説明したように、第3実施形態に係る機械学習装置は、追加データを表すトリプルから予測ノードまでのノード及びエッジが含まれるように範囲

を決定する。これにより、追加データによる影響を予測に及ぼすことができる。

[0065] なお、第3実施形態では、追加データを表すトリプルの一方のノードから予測ノードまでのパスを含めるように範囲を拡張する場合について説明したが、他方のノード側も範囲を拡張するようにしてもよい。具体的には、更新部が、追加データを表すトリプルの他方のノードから、一方のノードから予測ノードまでの距離と同一の距離となるノードまでに含まれるノード及びエッジが含まれるように範囲を決定してもよい。例えば、図18の例では、サブジェクトであるノードから予測ノードまでのパスが含まれるように範囲を決定しているが、このパスの長さ（距離に相当）は2である。したがって、更新部は、図19に示すように、オブジェクトであるノードから長さ2となるパスに含まれるエッジ及びノードも含まれるように範囲24を拡張してよい。これにより、追加データを表すトリプルのサブジェクトであるノード側とオブジェクトであるノード側とで、オンライン学習の対象となる範囲の不均衡を解消することができる。

[0066] なお、上記各実施形態では、1つのコンピュータで、機械学習処理と予測処理とを行う場合について説明したが、これに限定されない。取得部及び更新部を含む機械学習装置と、予測部を含む予測装置とを、それぞれ異なるコンピュータで実現してもよい。また、この場合、機械学習装置の記憶部には、上記各実施形態における機械学習処理を実行するための機械学習プログラムが記憶される。また、予測装置の記憶部には、上記各実施形態における予測処理を実行するための予測プログラムが記憶される。

[0067] また、上記各実施形態では、機械学習プログラムが記憶部に予め記憶（インストール）されている態様を説明したが、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

[0068] 10、210、310 機械学習装置

- 1 2 取得部
- 1 4、2 1 4、3 1 4 更新部
- 1 6 機械学習済みのグラフデータ
- 1 8 予測部
- 2 0 更新されたグラフデータ
- 2 2 追加データを表すトリプル
- 2 4 オンライン学習の範囲
- 2 4 A、2 4 B、2 4 C、2 4 D、2 4 E オンライン学習の範囲に含まれるトリプル
- 2 6 予測対象のエッジ
- 2 8 予測ノード
- 4 0 コンピュータ
- 4 1 C P U
- 4 2 メモリ
- 4 3 記憶部
- 4 9 記憶媒体
- 5 0、2 5 0、3 5 0 機械学習プログラム

請求の範囲

- [請求項1] 第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報を取得し、
- 前記情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルを更新する、
- 処理をコンピュータに実行させることを特徴とする機械学習プログラム。
- [請求項2] 前記特定の条件は、前記複数のノードと前記複数のエッジとを含むグラフデータに前記第1のトリプルを追加した場合に、前記第1のトリプルと前記一又は複数のノードの各々との間のエッジ数で表される前記第1のトリプルとの距離が所定値以下であることを含む、
- ことを特徴とする請求項1に記載の機械学習プログラム。
- [請求項3] 前記所定値が、エッジ数1で表される距離である、
- ことを特徴とする請求項2に記載の機械学習プログラム。
- [請求項4] エッジの始点側のノードをサブジェクト、エッジの終点側のノードをオブジェクトとする場合において、前記複数のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合が所定割合以下の場合、
- 前記特定の条件は、前記第1のトリプルにおいてサブジェクトであるノードをオブジェクトとして接続すること、及び前記第1のトリプルにおいてオブジェクトであるノードをサブジェクトとして接続することを含む、
- ことを特徴とする請求項3に記載の機械学習プログラム。

[請求項5] エッジの始点側のノードをサブジェクト、エッジの終点側のノードをオブジェクトとする場合において、前記複数のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合が所定割合を超える場合、

前記特定の条件は、前記第1のトリプルにおいてサブジェクトであるノードをサブジェクトとして接続すること、及び前記第1のトリプルにおいてオブジェクトであるノードをオブジェクトとして接続することを含む、

ことを特徴とする請求項3に記載の機械学習プログラム。

[請求項6] 前記特定の条件は、前記第1のノード及び前記第2のノードのうち、前記グラフデータを用いた予測の際に予測の対象となるエッジに接続する予測ノードまでの距離が短い一方のノードから前記予測ノードまでに含まれるノード及びエッジであることを含む、

ことを特徴とする請求項2～請求項5のいずれか1項に記載の機械学習プログラム。

[請求項7] 前記特定の条件は、前記第1のノード及び前記第2のノードの他方のノードから、前記一方のノードから前記予測ノードまでの距離と同一の距離となるノードまでに含まれるノード及びエッジであることを含む、

ことを特徴とする請求項6に記載の機械学習プログラム。

[請求項8] 予測対象の入力データを取得し、

第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルが更新された前

記複数のノードと前記複数のエッジとを含むグラフデータを用いて、前記入力データにおける、予測の対象となるエッジの有無を予測する処理をコンピュータに実行させることを特徴とする予測プログラム。

[請求項9] 第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報を取得し、

前記情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルを更新する、

処理を実行する制御部を含むことを特徴とする機械学習装置。

[請求項10] 前記特定の条件は、前記複数のノードと前記複数のエッジとを含むグラフデータに前記第1のトリプルを追加した場合に、前記第1のトリプルと前記一又は複数のノードの各々との間のエッジ数で表される前記第1のトリプルとの距離が所定値以下であることを含む、

ことを特徴とする請求項9に記載の機械学習装置。

[請求項11] 前記所定値が、エッジ数1で表される距離である、

ことを特徴とする請求項10に記載の機械学習装置。

[請求項12] エッジの始点側のノードをサブジェクト、エッジの終点側のノードをオブジェクトとする場合において、前記複数のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合が所定割合以下の場合、

前記特定の条件は、前記第1のトリプルにおいてサブジェクトであるノードをオブジェクトとして接続すること、及び前記第1のトリプルにおいてオブジェクトであるノードをサブジェクトとして接続する

ことを含む、

ことを特徴とする請求項 11 に記載の機械学習装置。

[請求項13]

エッジの始点側のノードをサブジェクト、エッジの終点側のノードをオブジェクトとする場合において、前記複数のノードのうち、サブジェクト及びオブジェクトの双方となるノードの数の割合が所定割合を超える場合、

前記特定の条件は、前記第1のトリプルにおいてサブジェクトであるノードをサブジェクトとして接続すること、及び前記第1のトリプルにおいてオブジェクトであるノードをオブジェクトとして接続することを含む、

ことを特徴とする請求項 11 に記載の機械学習装置。

[請求項14]

前記特定の条件は、前記第1のノード及び前記第2のノードのうち、前記グラフデータを用いた予測の際に予測の対象となるエッジに接続する予測ノードまでの距離が短い一方のノードから前記予測ノードまでに含まれるノード及びエッジであることを含む、

ことを特徴とする請求項 10～請求項 13 のいずれか 1 項に記載の機械学習装置。

[請求項15]

前記特定の条件は、前記第1のノード及び前記第2のノードの他方のノードから、前記一方のノードから前記予測ノードまでの距離と同一の距離となるノードまでに含まれるノード及びエッジであることを含む、

ことを特徴とする請求項 14 に記載の機械学習装置。

[請求項16]

予測対象の入力データを取得し、

第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトル

ルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルが更新された前記複数のノードと前記複数のエッジとを含むグラフデータを用いて、前記入力データにおける、予測の対象となるエッジの有無を予測する処理を実行する制御部を含むことを特徴とする予測装置。

[請求項17]

第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報を取得し、

前記情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルを更新する、

処理をコンピュータが実行することを特徴とする機械学習方法。

[請求項18]

予測対象の入力データを取得し、

第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルが更新された前記複数のノードと前記複数のエッジとを含むグラフデータを用いて、前記入力データにおける、予測の対象となるエッジの有無を予測する処理をコンピュータが実行することを特徴とする予測方法。

[請求項19]

第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報を取得

し、

前記情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルを更新する、

処理をコンピュータに実行させることを特徴とする機械学習プログラムを記憶した記憶媒体。

[請求項20]

予測対象の入力データを取得し、

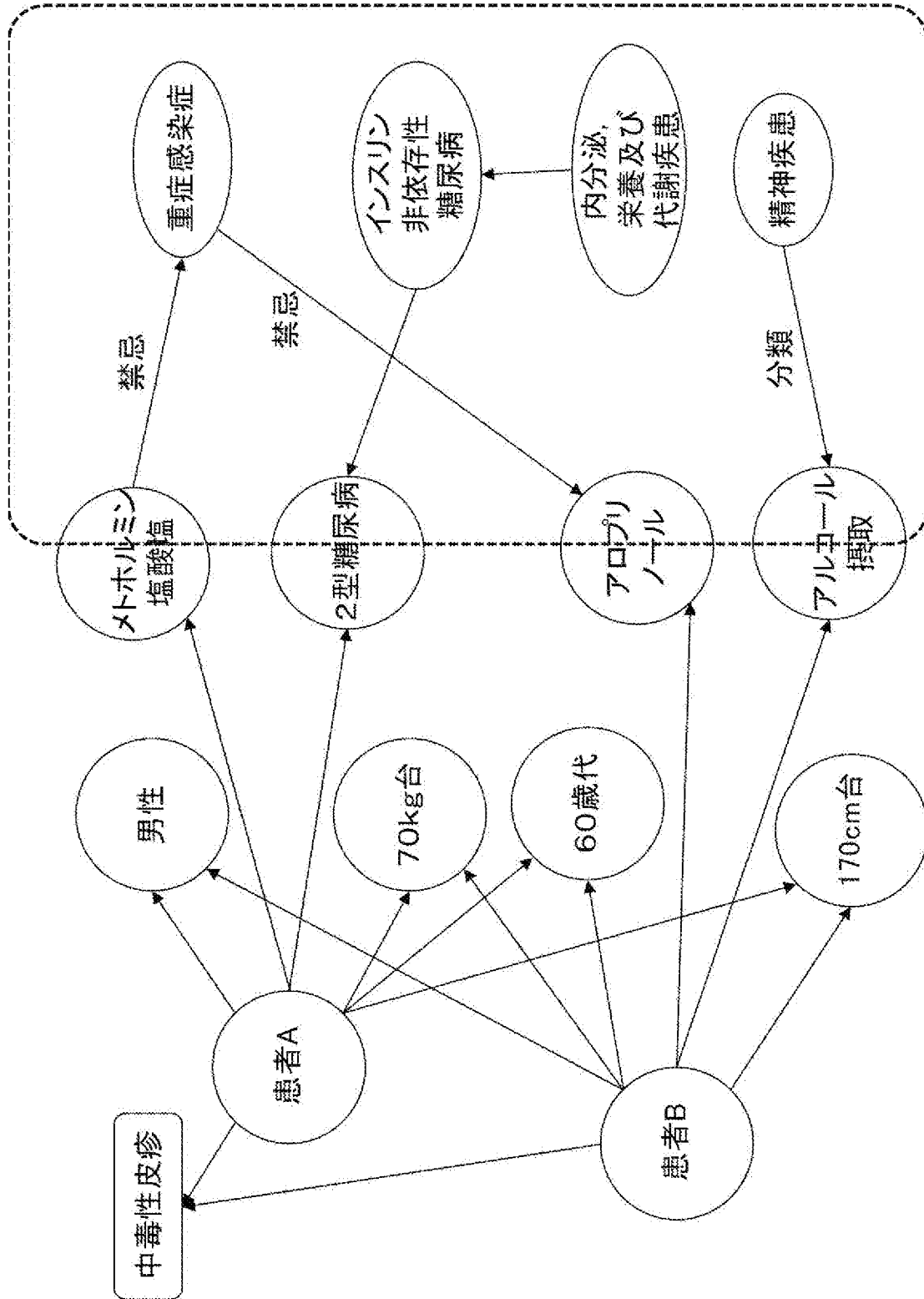
第1のノードと第2のノードと前記第1のノードと前記第2のノードとの関係を示す第1のエッジとを含む第1のトリプルの情報と、複数のノードと前記複数のノード間の関係を示す複数のエッジとを用いた機械学習によって生成された前記複数のノードと前記複数のエッジとのそれぞれを表す複数のベクトルとに基づいて、前記複数のベクトルのうち前記第1のトリプルと特定の条件で接続される一又は複数のノードと一又は複数のエッジとのそれぞれのベクトルが更新された前記複数のノードと前記複数のエッジとを含むグラフデータを用いて、前記入力データにおける、予測の対象となるエッジの有無を予測する

処理をコンピュータに実行させることを特徴とする予測プログラムを記憶した記憶媒体。

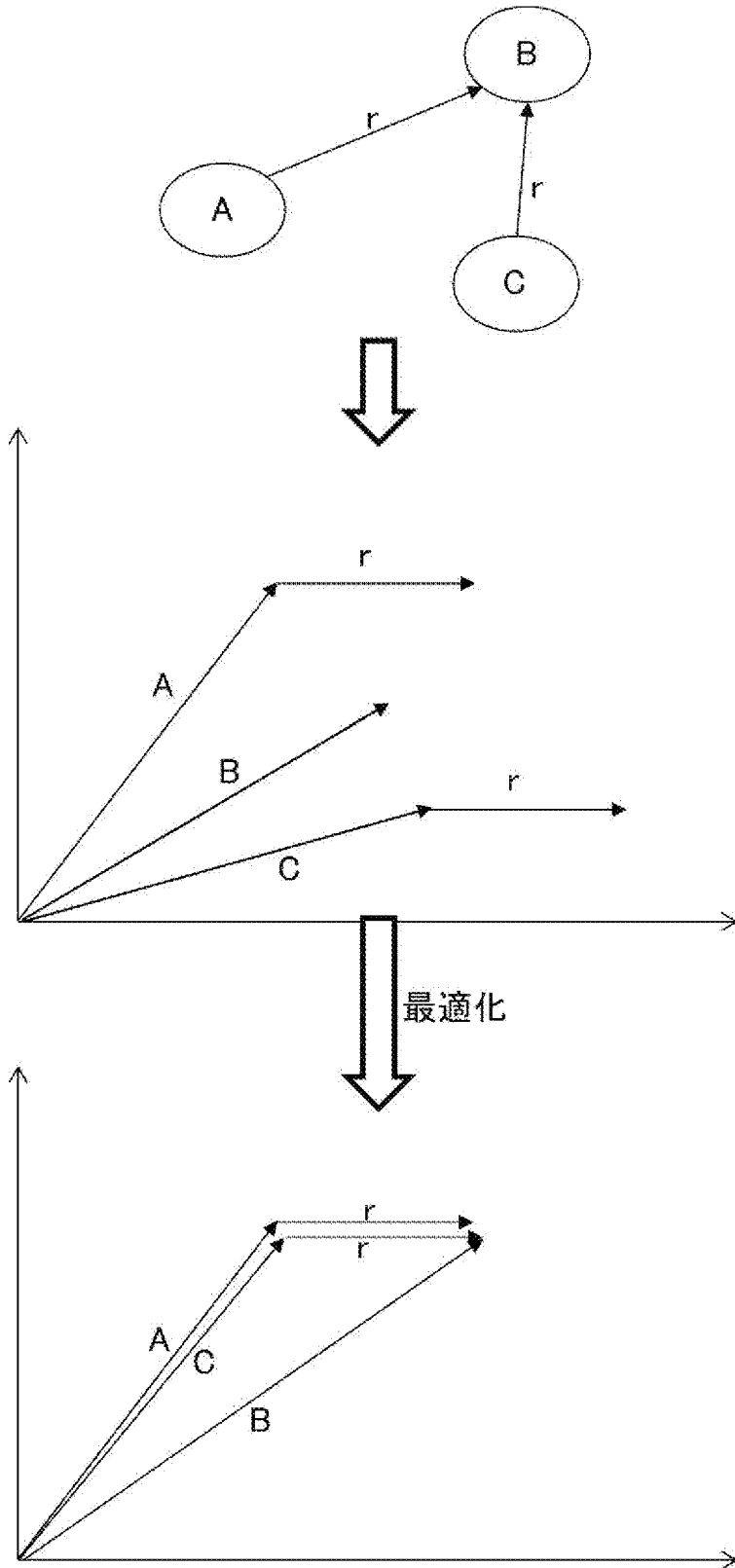
[図1]

ID	性別	年代	体重	身長	医薬品	疾病	副作用
A	男性	60歳代	70Kg台	170cm台	メトホルミン塩酸塩	2型糖尿病	中毒性皮疹
B	男性	60歳代	70Kg台	170cm台	アロブリンノール	アルコール摂取	中毒性皮疹
...							

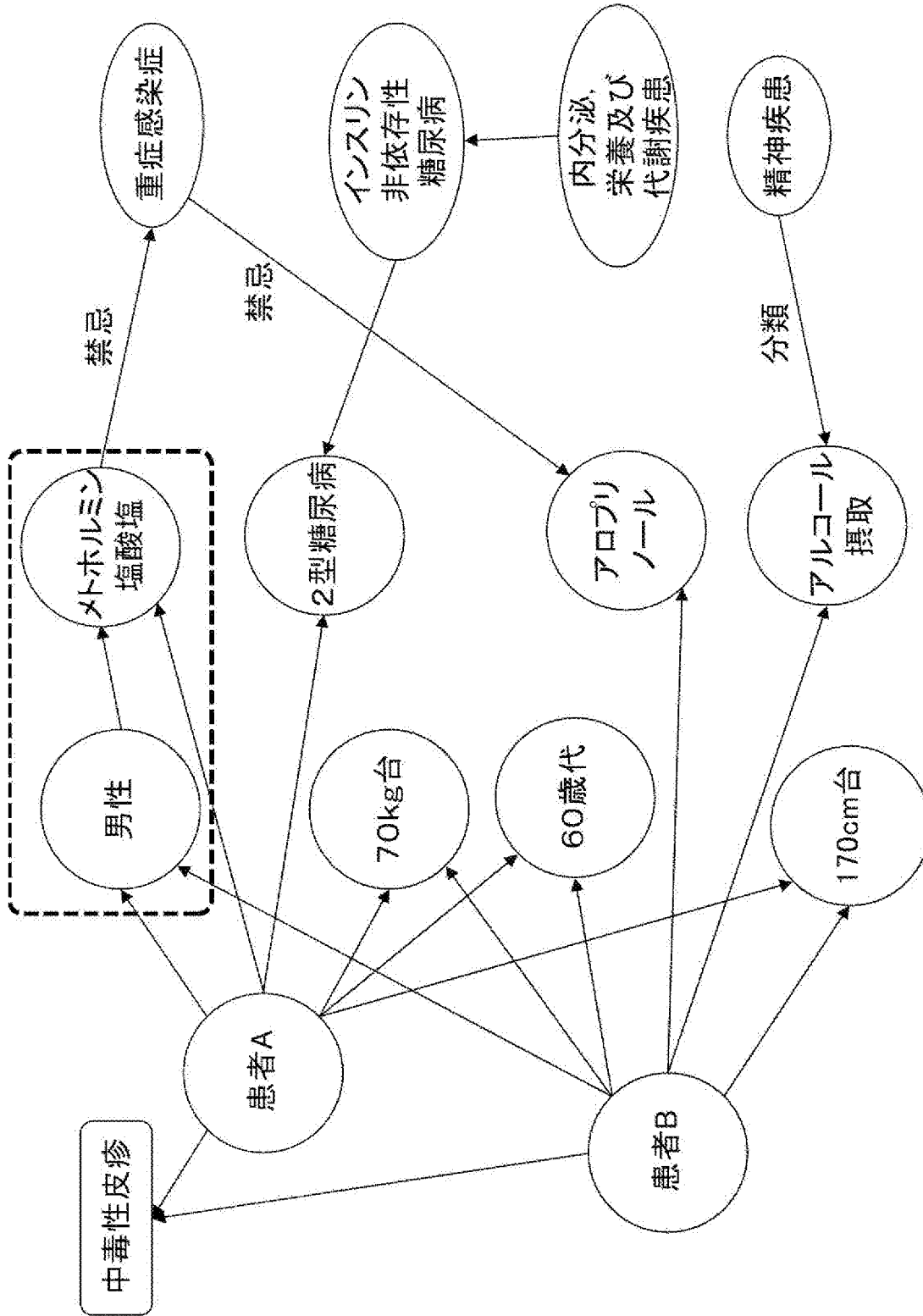
[図2]



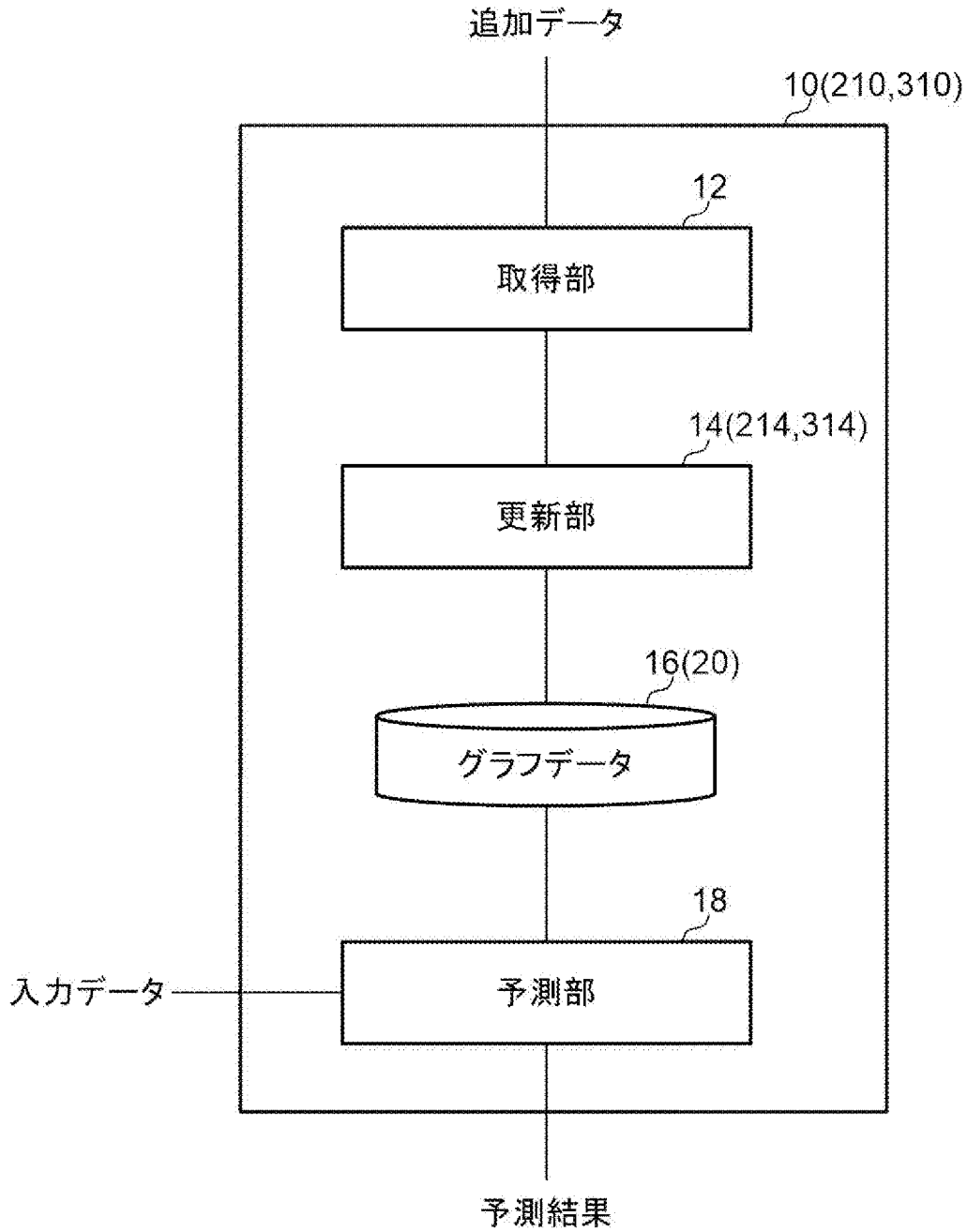
[図3]



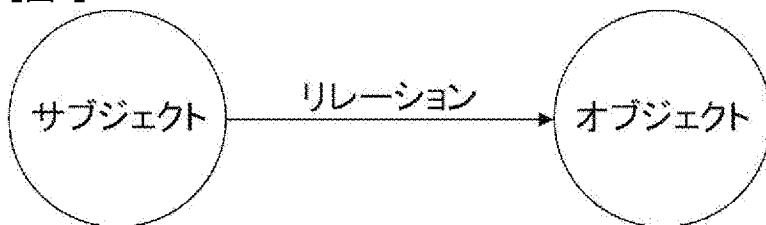
[図4]



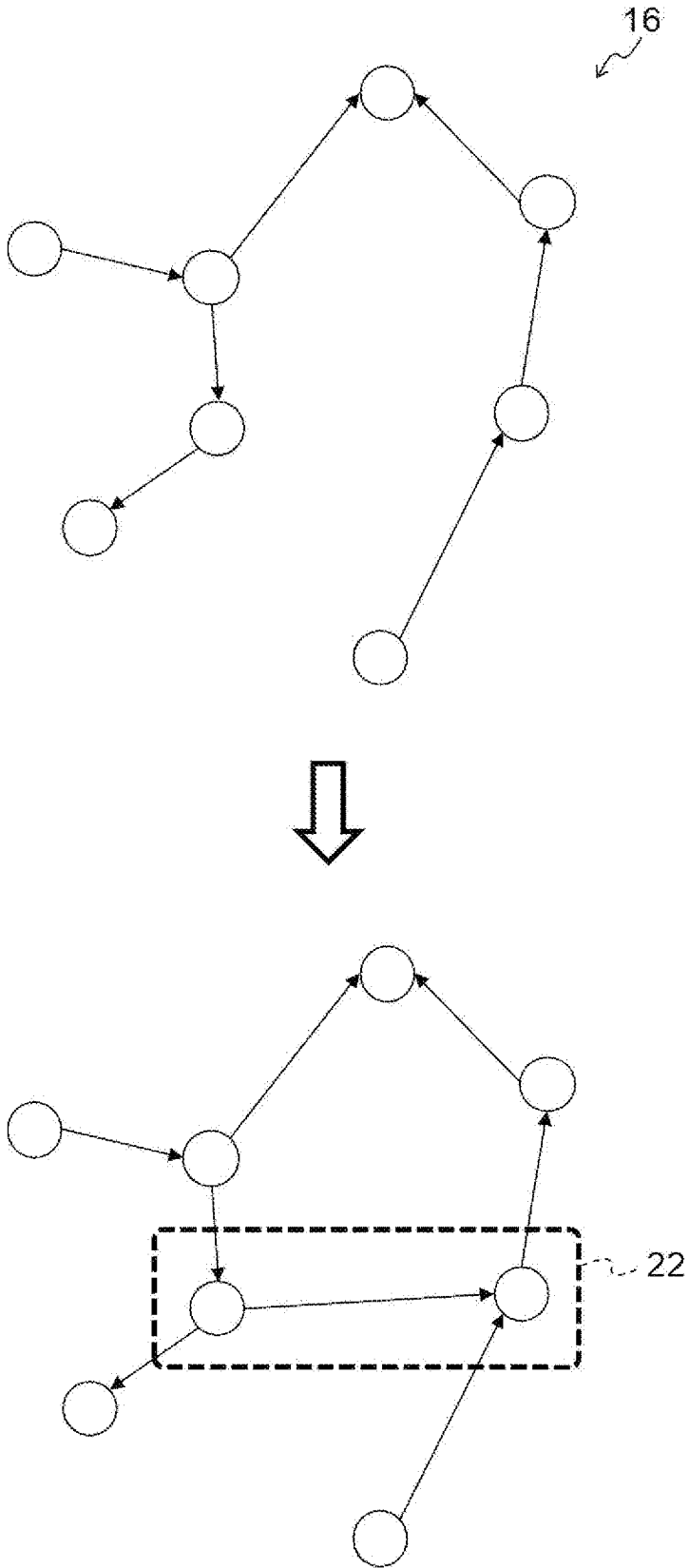
[図5]



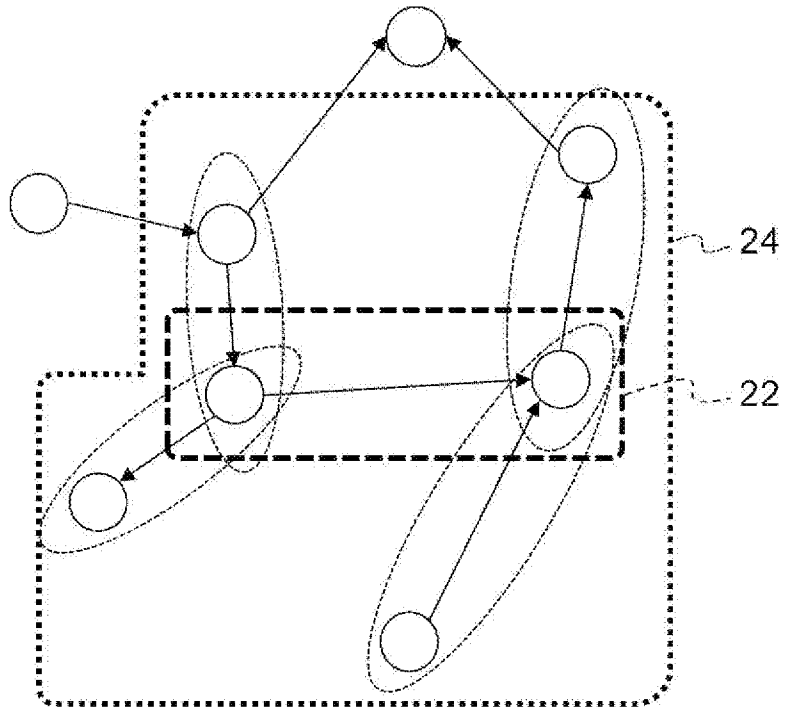
[図6]



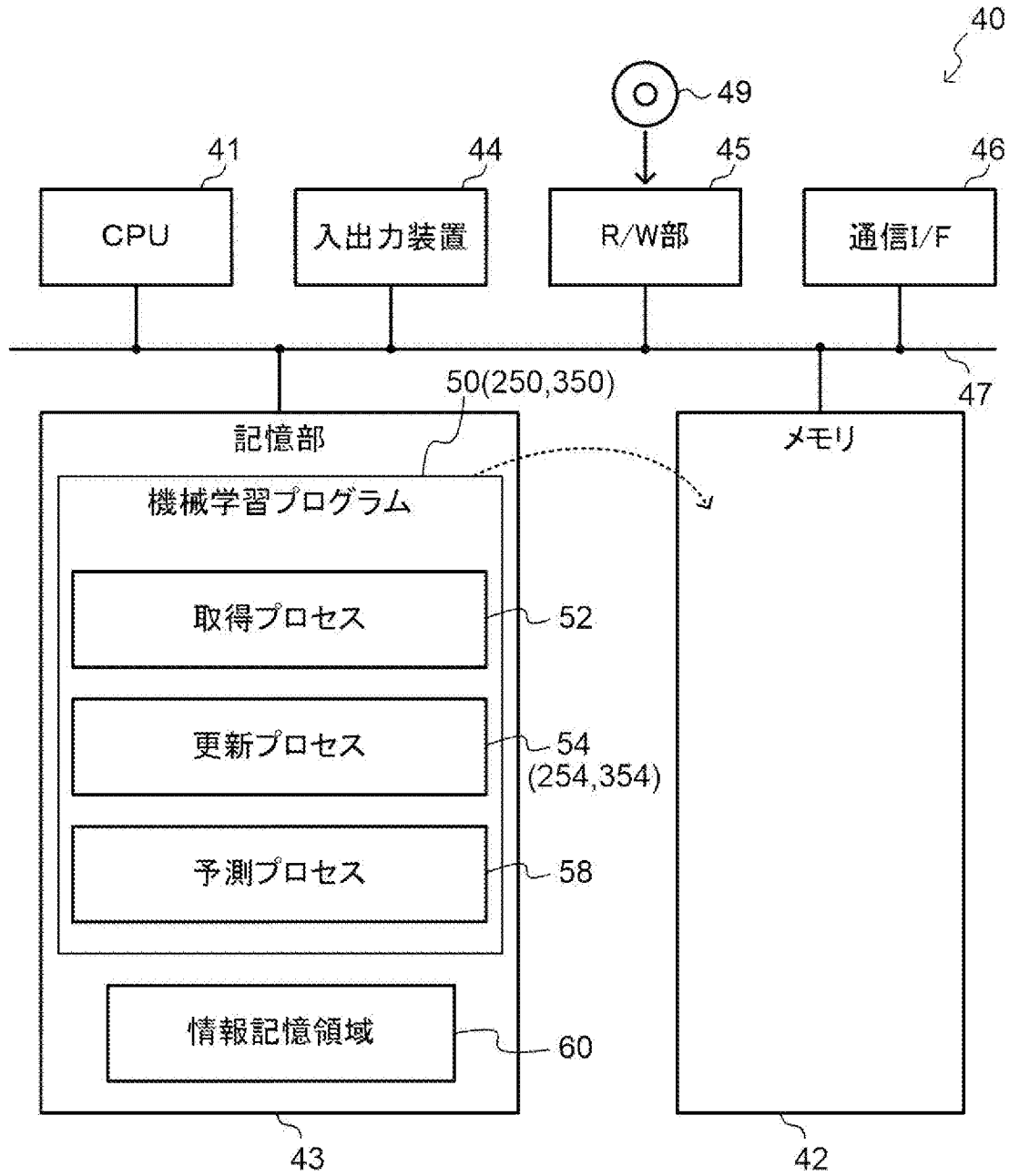
[図7]



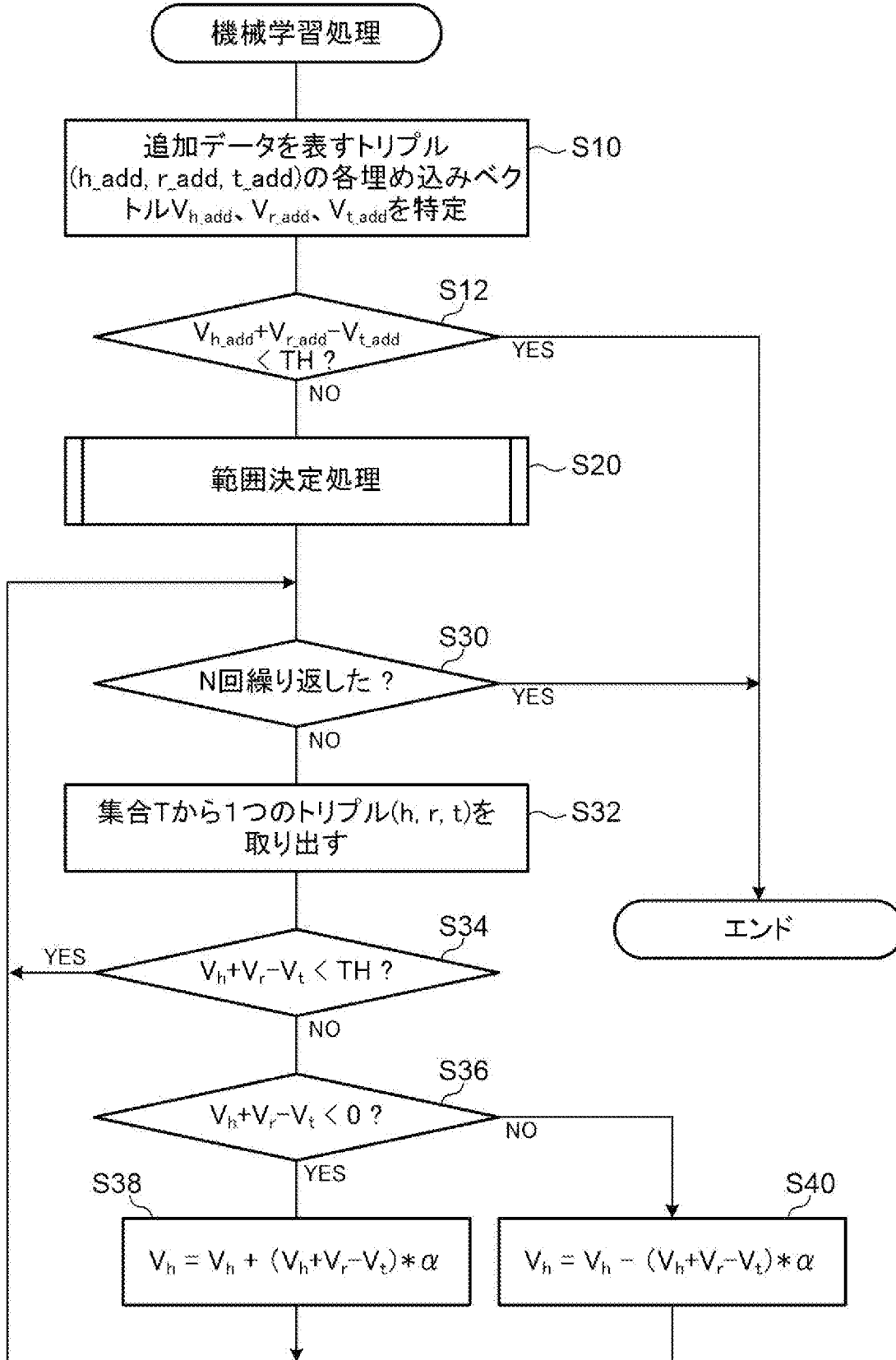
[図8]



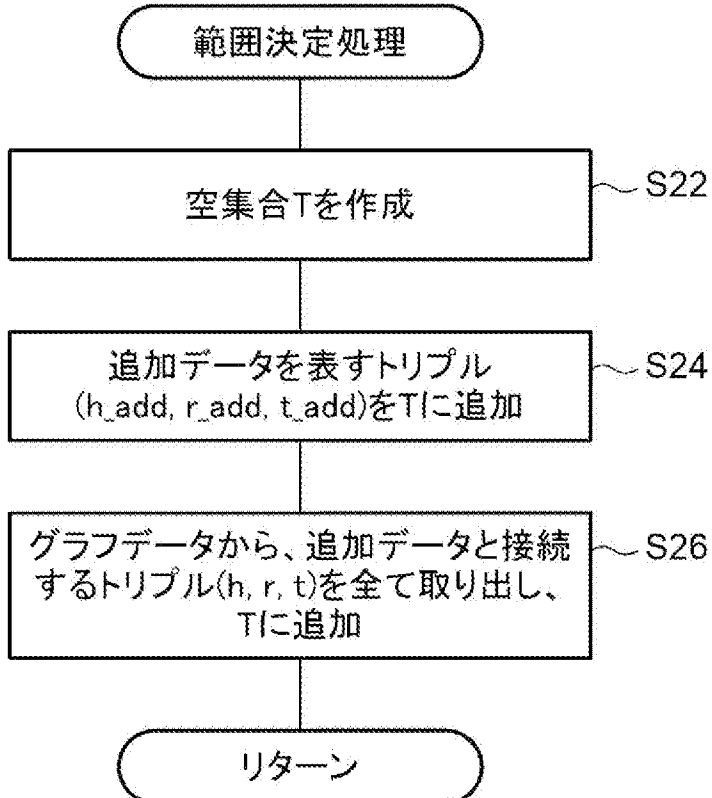
[図9]



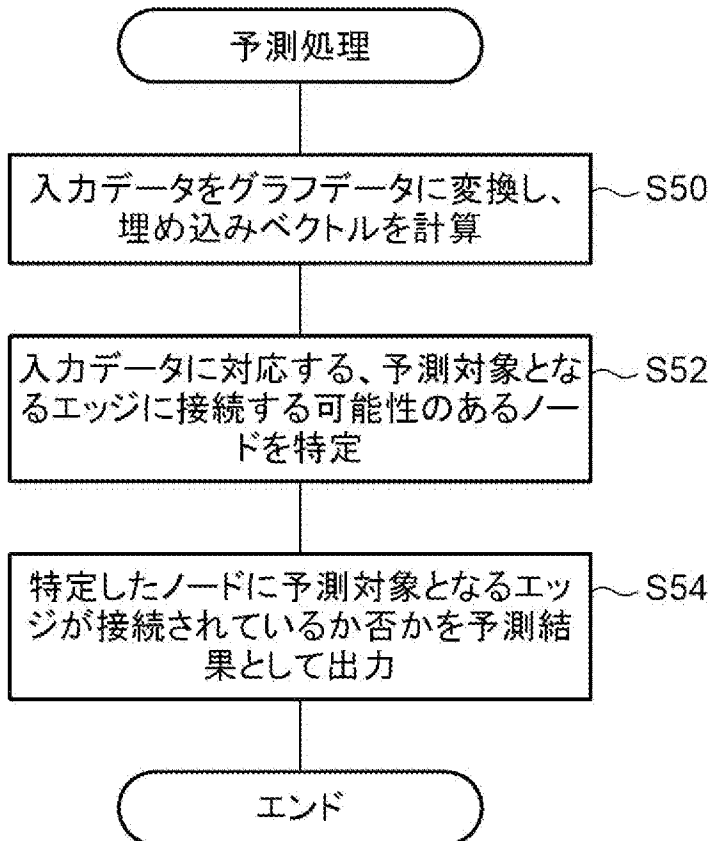
[図10]



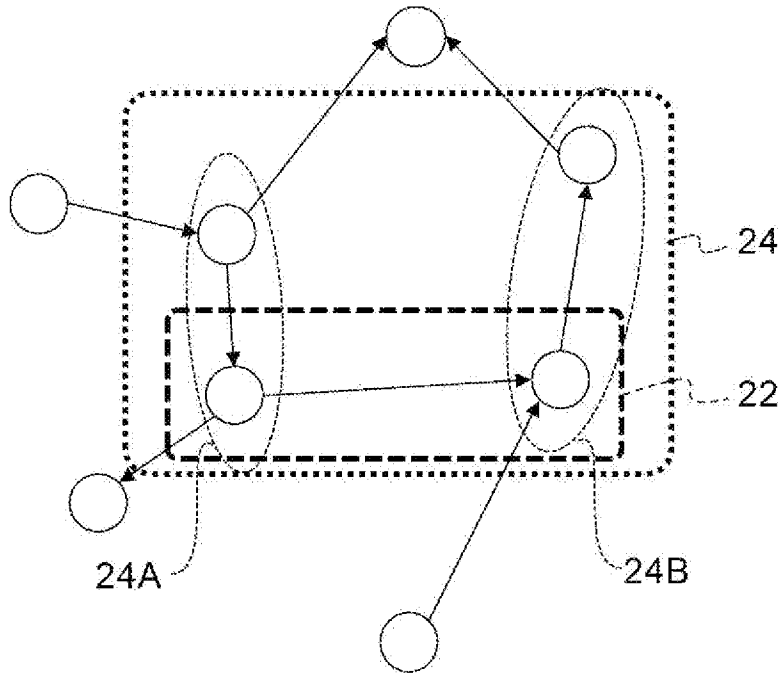
[図11]



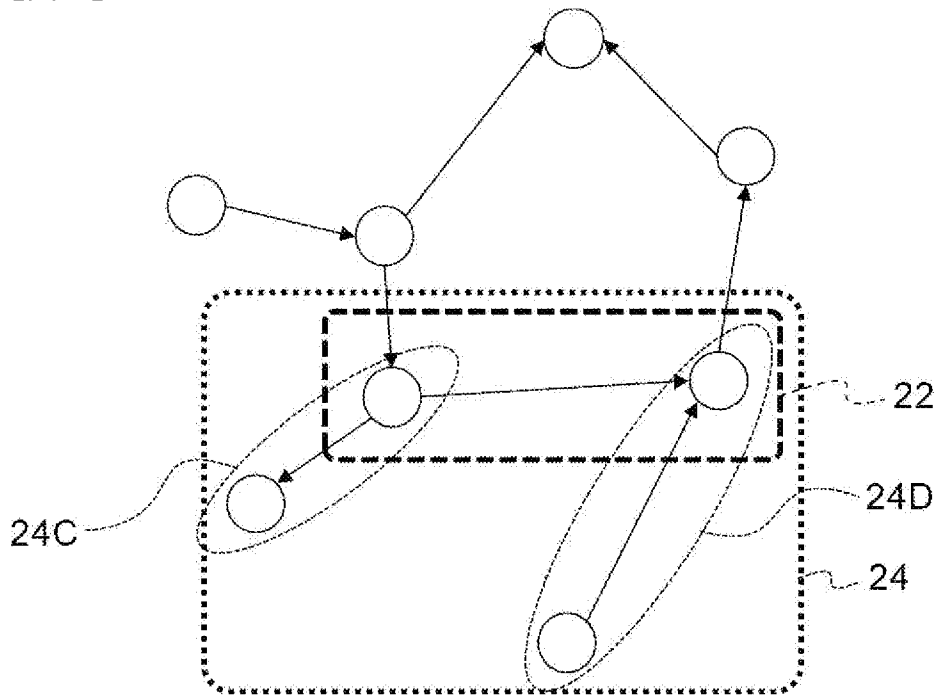
[図12]



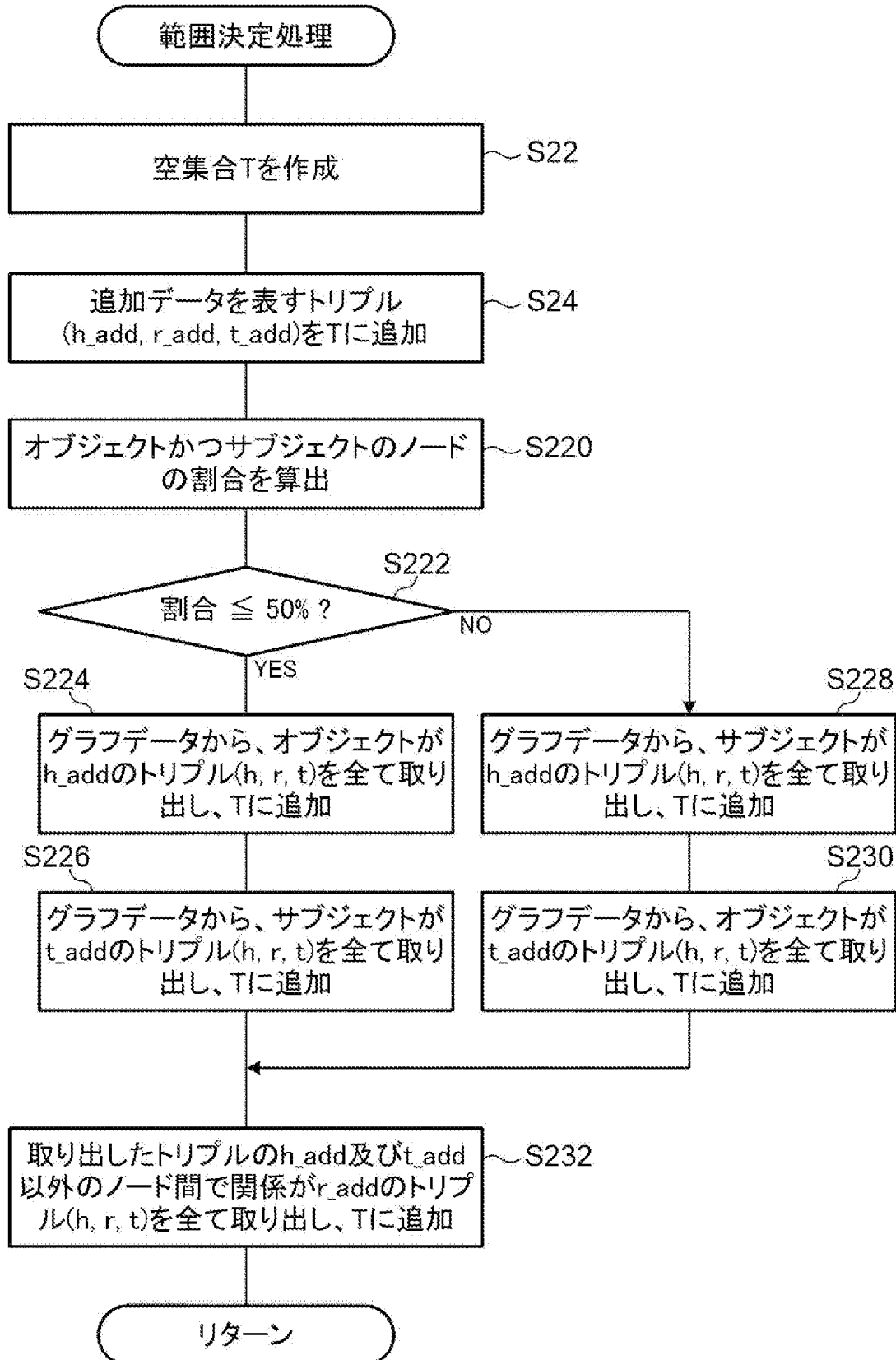
[図13]



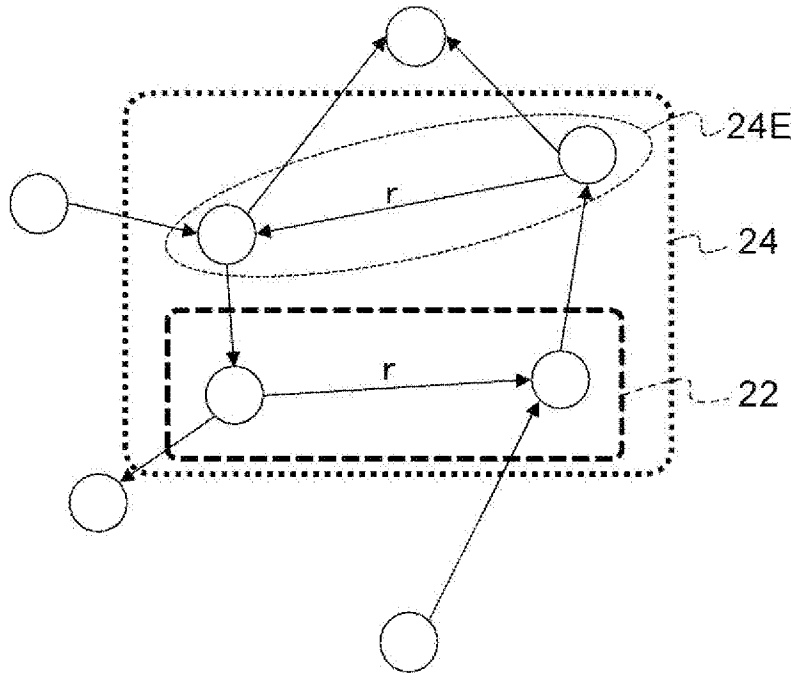
[図14]



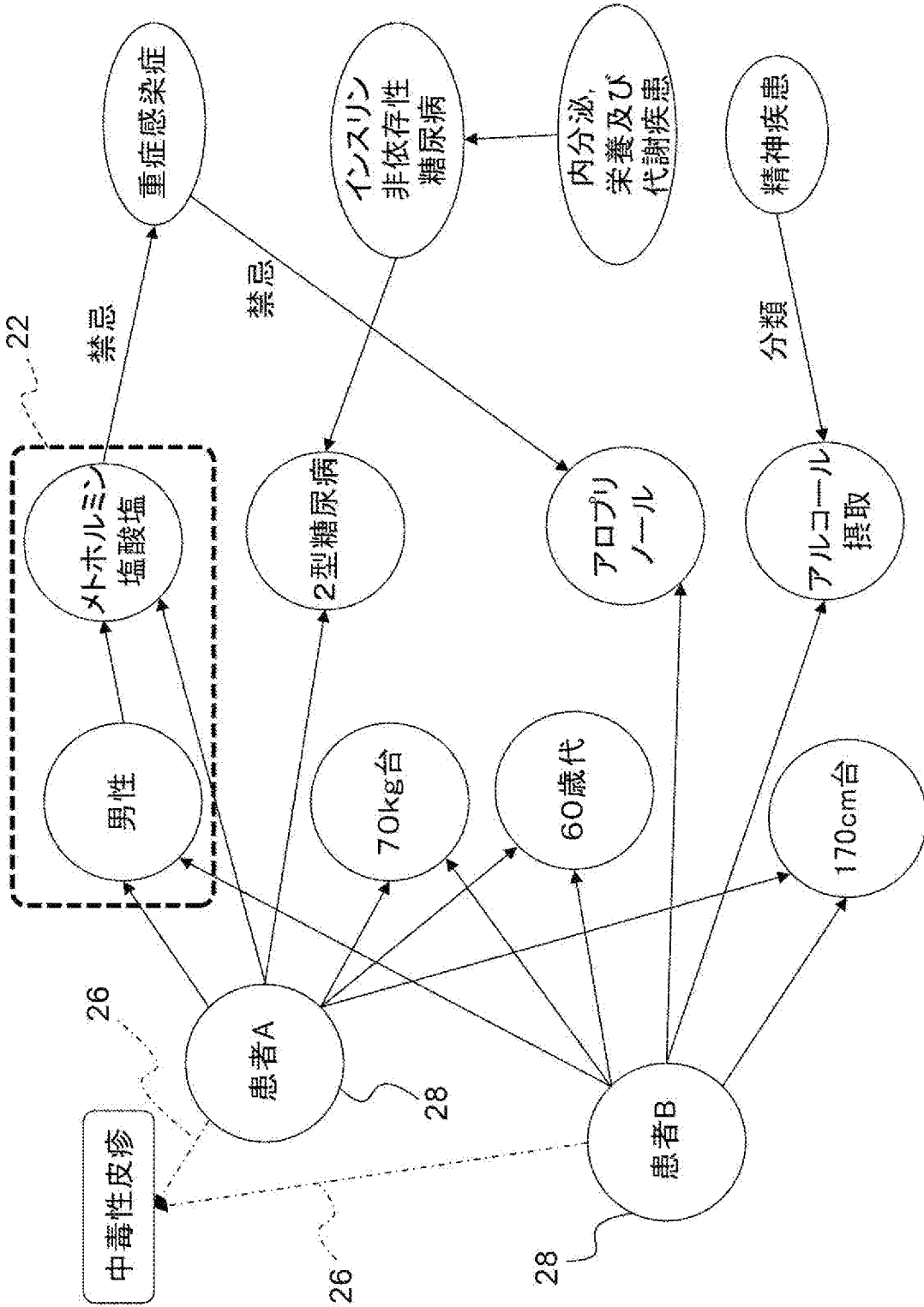
[図15]



[図16]



[図17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/009344

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06N20/00 (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)

Int. Cl. G06N20/00

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 1971-2021
Registered utility model specifications of Japan 1996-2021
Published registered utility model applications of Japan 1994-2021

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.
X	村上勝彦. ネットワーク構造を利用した知識グラフの埋込. 一般社団法人人工知能学会第34回全国大会(2020). [online], 09 June 2020, pp. 1-3 (3Rin4-31), in particular, see "2.3. Mediating Entity Key Point Learning", "4. Effect of Method", (MURAKAMI, Katsuhiko. Knowledge graph embedding using network structure. The 34th Annual Conference for the Japanese Society of Artificial Intelligence.)	1, 9, 17, 19 2-8, 10-16, 18, 20
A		8, 16, 18, 20 1-7, 9-15, 17, 19
X	US 2019/0220524 A1 (ACCENTURE GLOBAL SOLUTIONS LIMITED) 18 July 2019 (2019-07-18), paragraphs [0011], [0012], [0015], [0020], [0021], [0031]-[0034], [0057]	1-20
A		1-20

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

"I" 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
17.05.2021

Date of mailing of the international search report
25.05.2021

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/009344

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-085116 A (FUJITSU LTD.) 31 May 2018 (2018-05-31), paragraphs [0063]-[0066]	1-20
A	JP 2016-212853 A (FUJITSU LTD.) 15 December 2016 (2016-12-15), paragraphs [0014]-[0031]	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/009344

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
US 2019/0220524 A1	18.07.2019	US 10157226 B1	
JP 2018-085116 A	31.05.2018	US 2018/0144252 A1 paragraphs [0069]- [0072] EP 3327582 A1 DE 102016223193 A1	
JP 2016-212853 A	15.12.2016	US 2016/0321407 A1 paragraphs [0014]- [0031] GB 2537925 A EP 3089060 A1	

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）） G06N20/00 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 1971-2021年 日本国実用新案登録公報 1996-2021年 日本国登録実用新案公報 1994-2021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
X	村上勝彦, ネットワーク構造を利用した知識グラフの埋込, 一般社団法人 人工知能学会 第34回全国大会 (2020) [online], 2020.06.09, pp. 1-3 (3Rin4-31) 特に「2.3 媒介的エンティティの重点学習」及び「4. 本手法の効果」を参照。	1, 9, 17, 19
A		2-8, 10-16, 18, 20
X	US 2019/0220524 A1 (ACCENTURE GLOBAL SOLUTIONS LIMITED) 18.07.2019 (2019-07-18) 段落[0011]-[0012], [0015], [0020]-[0021], [0031]-[0034]及び[0057]	8, 16, 18, 20
A		1-7, 9-15, 17, 19
A	JP 2018-085116 A (富士通株式会社) 31.05.2018 (2018-05-31) 段落[0063]-[0066]	1-20
A	JP 2016-212853 A (富士通株式会社) 15.12.2016 (2016-12-15) 段落[0014]-[0031]	1-20
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" 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 17.05.2021	国際調査報告の発送日 25.05.2021	
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） 坂庭 剛史 5B 9288 電話番号 03-3581-1101 内線 3545	

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/JP2021/009344

引用文献	公表日	パテントファミリー文献	公表日
US 2019/0220524 A1	18.07.2019	US 10157226 B1	
JP 2018-085116 A	31.05.2018	US 2018/0144252 A1 段落[0069]-[0072]	
		EP 3327582 A1	
		DE 102016223193 A1	
JP 2016-212853 A	15.12.2016	US 2016/0321407 A1 段落[0014]-[0031]	
		GB 2537925 A	
		EP 3089060 A1	