

(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
2024年12月12日(12.12.2024)



(10) 国際公開番号

WO 2024/252480 A1

(51) 国際特許分類:
G06F 16/21 (2019.01) G06F 40/16 (2020.01)

(21) 国際出願番号: PCT/JP2023/020846

(22) 国際出願日: 2023年6月5日(05.06.2023)

(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: 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: 倉林 利行 (KURABAYASHI, Toshiyuki); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 1 N T T 知的財産センタ内 Tokyo (JP).

丹野 治門 (TANNO, Haruto); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 1 N T T 知的財産センタ内 Tokyo (JP).

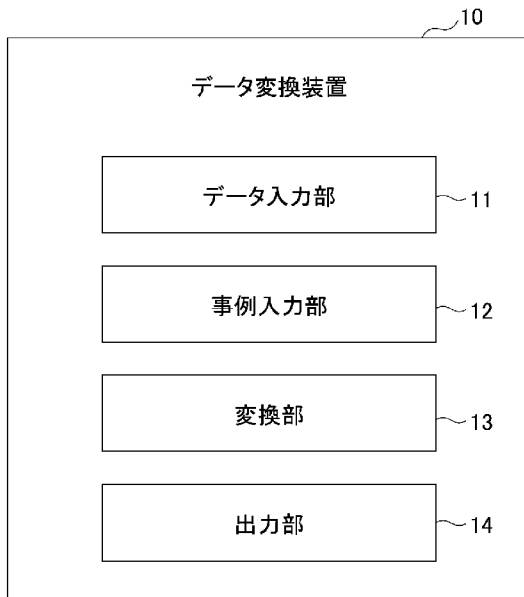
(74) 代理人: 伊東 忠重, 外 (ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,

(54) Title: DATA CONVERSION DEVICE, DATA CONVERSION METHOD, AND PROGRAM

(54) 発明の名称: データ変換装置、データ変換方法及びプログラム

[図5]



10 Data conversion device
11 Data input unit
12 Example input unit
13 Conversion unit
14 Output unit

(57) Abstract: A data conversion device (10) comprises: a conversion target table; an example input unit (12) for inputting a conversion result of a positive example and a conversion result of a negative example related to a portion of the table; a conversion unit (13) that generates a table which includes the positive example and does not include the negative example by setting the conversion target table as the initial operation target and then repeating the process of calculating an evaluation value based on a comparison between the positive example and the negative example for each table obtained by ex-

[続葉有]

LY, MA, MD, MG, MK, MN, MU, 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, CV, 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, ME, 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))

ecuting, among a plurality of operations, an operation that can be executed on the operation target, and setting a table with the best evaluation value as the operation target; and an output unit (14) that outputs the table which includes the positive example and does not include the negative example. If a new positive example or a new negative example is input by the input unit (12) after the output unit (14) outputs a table, the conversion unit (13) generates a table which includes the new positive example and does not include the new negative example, thereby improving work efficiency for data conversion.

(57) 要約：データ変換装置（10）は、変換対象のテーブルと、前記テーブルの一部に関する正例の変換結果と負例の変換結果とを入力する事例入力部（12）と、前記変換対象のテーブルを最初の操作対象とし、複数の操作のうち前記操作対象に対して実行可能な操作を実行することで得られるテーブルごとに、前記正例及び前記負例との比較に基づく評価値を計算し、前記評価値が最良であるテーブルを前記操作対象とすることを繰り返すことで、前記正例を含み前記負例を含まないテーブルを生成する変換部（13）と、前記正例を含み前記負例を含まないテーブルを出力する出力部（14）と、を有し、前記変換部（13）は、前記出力部（14）がテーブルを出力した後に、前記事例入力部（12）が新たな正例又は新たな負例を入力した場合に、前記新たな正例をも含み前記新たな負例をも含まないテーブルを生成することで、データ変換のための作業効率を向上させる。

明 細 書

発明の名称：データ変換装置、データ変換方法及びプログラム
技術分野

[0001] 本発明は、データ変換装置、データ変換方法及びプログラムに関する。

背景技術

[0002] デジタル・トランスフォーメーションによって、デジタル技術とデータ（情報資産）の活用が進みつつある。より競争力を高めるためにはデータサイエンティストだけでなく、一般社員も日常業務にデータ分析を組み込んでいくことが重要である。データ分析を行うためにはデータの前処理が必要となるが、前処理には約8割の稼働がかかるとも言われており、変換方法やプログラミングの知識も要するため、データ分析の普及の妨げとなっている。

[0003] データ分析を支援するツールとして、ETLツールが存在する。ETLツールではデータの変換方法を指定することで、データ変換を自動で行うことができる。しかし、変換方法の知識が必要なことに加え、新しいデータ変換を行うたびに変換方法を指定するための手間がかかる。

[0004] AutoPandas（非特許文献1）は、データ変換前後の例から、その変換を実現するプログラムを合成する技術である。AutoPandasではデータ変換前後の具体例を用意するだけで良いため、変換方法やプログラミングの知識がなくても使うことができるという利点が存在する。

先行技術文献

非特許文献

[0005] 非特許文献1：Rohan Bavishi, Caroline Lemieux, Roy Fox, Koushik Sen, and Ion Stoica. 2019. "AutoPandas: neural-backed generators for program synthesis." Proc. ACM Program. Lang. 3, OOPSLA, Article 168 (October 2019), 27 pages

発明の概要

発明が解決しようとする課題

[0006] しかし、ユーザは実現したい仕様を反映したデータ変換前後の例を用意する必要があり、データ変換前後の例に対する仕様の反映が不十分で所望のプログラムが合成できなかった場合、再度その例を作成し直さなくてはならず、多くの稼働がかかる。

[0007] 本発明は、上記の点に鑑みてなされたものであって、データ変換のための作業効率を向上させることを目的とする。

課題を解決するための手段

[0008] そこで上記課題を解決するため、データ変換装置は、変換対象のテーブルと、前記テーブルの一部に関する正例の変換結果と負例の変換結果とを入力するように構成されている事例入力部と、前記変換対象のテーブルを最初の操作対象とし、複数の操作のうち前記操作対象に対して実行可能な操作を実行することで得られるテーブルごとに、前記正例及び前記負例との比較に基づく評価値を計算し、前記評価値が最良であるテーブルを前記操作対象とすることを繰り返すことで、前記正例を含み前記負例を含まないテーブルを生成するように構成されている変換部と、前記正例を含み前記負例を含まないテーブルを出力するように構成されている出力部と、を有し、前記変換部は、前記出力部がテーブルを出力した後に、前記事例入力部が新たな正例又は新たな負例を入力した場合に、前記新たな正例をも含み前記新たな負例をも含まないテーブルを生成するように構成されている。

発明の効果

[0009] データ変換のための作業効率を向上させることができる。

図面の簡単な説明

[0010] [図1]第1の実施の形態においてユーザが所望するデータ変換を説明するための図である。

[図2]正例及び負例の一例を示す図である。

[図3]探索グラフの一例を示す図である。

[図4]第1の実施の形態におけるデータ変換装置10のハードウェア構成例を示す図である。

[図5]第1の実施の形態におけるデータ変換装置10の機能構成例を示す図である。

[図6]データ変換装置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7]第1の実施の形態におけるデータ変換処理の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8]ノードに対応するテーブルに対するオペレータの実行処理の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9]負例の追加により生成される探索グラフの一例を示す図である。

[図10]第2の実施の形態におけるデータ変換処理の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11]各ノードのコストの再計算処理の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12]コストの再計算後の探索グラフGの一例を示す図である。

発明を実施するための形態

[0011] 以下、図面に基づいて本発明の実施の形態を説明する。第1の実施の形態の概要について説明する。第1の実施の形態では、ユーザと機械が対話的にデータ変換を行うことで、変換方法やプログラミングの知識がなくても、少ない稼働でデータ変換を行う手法が開示される。

[0012] 図1は、第1の実施の形態においてユーザが所望するデータ変換を説明するための図である。第1の実施の形態では、テーブル T_i についてテーブル T_o への変換をユーザが所望しているとする。なお、テーブル T_o は、ユーザの頭の中で描かれているにすぎず、実際には得られていないテーブルである。また、ユーザの頭の中では、全ての行及び全ての列の値が明確に特定されているわけではなく、どのような変換結果が欲しいのかについての漠然としたイメージが有るにすぎない。

[0013] まず、ユーザは変換後のデータの一部（一部のデータ）の正しい変換結果の事例（以下、「正例」という。）を機械（後述のデータ変換装置10）に

与える。例えば、図2の正例E_pが機械に与えられる。正例E_pは、テーブルT_oに含まれるデータの集合である。図1では正例E_pが1つの事例のみを含む例が示されているが、正例E_pは複数の事例を含みうる。機械は与えられた正例E_pが含まれるような変換プログラム（変換手順）を探索（ヒューリスティック探索）し、変換後のデータをユーザに提示する。ユーザはそのデータをレビューし、不適切な箇所があれば修正する。具体的には、ユーザは、不適切な箇所に該当するデータを誤った変換結果の事例（以下、「負例」という。）として与える。例えば、図2の負例E_nが機械に与えられる。負例E_nは、テーブルT_oに含まれないデータの集合である。図1では負例E_nが1つの事例のみを含む例が示されているが、負例E_nは複数の事例を含みうる。なお、更なる正例がE_pに追加されてもよい。機械はその修正を考慮した変換プログラム（与えられた全ての正例を含み与えられた全ての負例を含まない変換プログラム）を再度探索し、新しい変換後のデータを再度ユーザに提示する。このような対話的なプロセスを繰り返すことで、ユーザは変換方法やプログラミングの知識がなくても所望のデータ変換を行うことができる。

[0014] なお、変換プログラムは、テーブルに対する以下の操作（以下、「オペレータ」という。）の再帰的な繰り返しである。

- ・ `move__end (n)` : n列目の列を一番右の列へと移動する。但し、既に一番右となっている列には適用されない。
- ・ `split (n, str)` : n列目のデータを `str` で分割する。但し `str` が存在しない列には適用されない。
- ・ `del__empty ()` : 空欄が存在する行を削除する。但しテーブルに空欄が存在しない場合は適用されない。

[0015] 本実施の形態において変換プログラム（変換後のデータ）の探索には、探索グラフが用いられる。図3は、探索グラフの一例を示す図である。探索グラフは、ルートノードが変換対象であるテーブルT_iに対応し、エッジがオペレータに対応し、各ノードは親ノードに対して親ノードとの間のエッジに

係るオペレータを実行した結果として得られるテーブルに対応する有向グラフである。各ノードには、当該ノードに対応するテーブルがどの程度正例を含み、どの程度負例を含まないかについての定量的な評価結果である評価値（以下、「コスト」という。）が付与される。本実施の形態において、与えられた全ての正例を含み、与えられた全ての負例を含まないテーブルのコストは0である。すなわち、コストは0以上の値であり、0が最良である。

[0016] 図3の例では、テーブル T_i に対して実行可能なオペレータである`move__end(0)`及び`split(1, ', ')`が実行された状態における探索グラフが示されている。したがって、ルートノードからは、`move__end(0)`に対応するエッジと`split(1, ', ')`に対応するエッジとが伸びている。`move__end(0)`に対応するエッジの先の子ノードは、テーブル T_i に対して`move__end(0)`を実行することで得られるテーブル t_1 に対応する。`split(1, ', ')`に対応するエッジの先の子ノードは、テーブル T_i に対して`split(1, ', ')`を実行することで得られるテーブル t_2 に対応する。

[0017] 機械は、探索グラフの末端のノードの中でコストが最良であるノードに対してオペレータを実行することを再帰的に繰り返す。機械は、テーブル T_o に一致するノードが得られた時点で探索を終了する。図3における各テーブルのコストは、図1の正例 E_p のみが与えられている場合のコストである。しテーブル t_2 は、正例 E_p を含むためコストが0となっている。この場合、テーブル t_2 がユーザに対して提示される。ユーザは、テーブル t_2 において、テーブル T_o と異なる部分を除くため、負例 E_n を追加する。その後、機械は改めて負例 E_n をも満たす変換プログラム（変換後のデータ）を探索する。

[0018] 以下、このような機械として機能するデータ変換装置10について説明する。図4は、第1の実施の形態におけるデータ変換装置10のハードウェア構成例を示す図である。図4のデータ変換装置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、インタフェース装置１０５、表示装置１０６、及び入力装置１０７等を有する。

[0019] データ変換装置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。

[0020] 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データ変換装置１０に係る機能を実現する。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。表示装置１０６はプログラムによるＧＵＩ（Graphical User Interface）等を表示する。入力装置１０７はキーボード及びマウス等で構成され、様々な操作指示を入力させるために用いられる。

[0021] 図５は、第１の実施の形態におけるデータ変換装置１０の機能構成例を示す図である。図５において、データ変換装置１０は、データ入力部１１、事例入力部１２、変換部１３及び出力部１４を有する。これら各部は、データ変換装置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。

[0022] 以下、データ変換装置１０が実行する処理手順について説明する。図６は、データ変換装置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。

[0023] ステップＳ１０１において、データ入力部１１は、変換対象のテーブルを入力する。例えば、図１のテーブルＴｉが入力される。

[0024] 続いて、事例入力部１２は、変換に関する事例（正例又は負例）の入力を

ユーザから受け付ける（S 1 0 2）。例えば、最初は、図 2 の正例 E p が入力される。但し、当初から負例 E n も入力されてもよい。

[0025] 続いて、変換部 1 3 は、データ変換処理を実行する（S 1 0 3）。データ変換処理では、探索グラフを用いてコストが 0 であるテーブルが探索されるまで、テーブル T i に対して再帰的にオペレータが実行される。つまり、オペレータの実行対象（操作対象）の系列の中で末端のテーブルの中でコストが最良であるテーブルに対してオペレータを実行することが再帰的に繰り返される。変換部 1 3 は、コストが 0 であるテーブルが探索されると、当該テーブルを変換後データとして返却する。

[0026] 続いて、出力部 1 4 は、変換部 1 3 が生成した変換後データを出力する（S 1 0 4）。

[0027] 続いて、事例入力部 1 2 は、出力された変換後データの適否をユーザから受け付ける（S 1 0 5）。適切である旨の入力を受け付けた場合（S 1 0 5 で Y e s）、図 6 の処理手順は終了する。この場合、ステップ S 1 0 4 において出力された変換後データが、ユーザによって利用される。

[0028] 一方、適切でない旨の入力を受け付けた場合（S 1 0 5 で N o）、ステップ S 1 0 2 以降が繰り返される。2 回目以降に実行されるステップ S 1 0 2 では、新たな正例又は新たな負例が追加される。

[0029] 続いて、ステップ S 1 0 3 の詳細について説明する。図 7 は、第 1 の実施の形態におけるデータ変換処理の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。

[0030] ステップ S 2 1 0 において、変換部 1 3 は、探索グラフ G を空集合（ Φ ）で初期化する。続くステップ S 2 1 1 ~ S 2 1 4 では探索グラフ G に対してルートノードが生成される。

[0031] ステップ S 2 1 1 において、変換部 1 3 は、parent に null を代入し、edge に null を代入し、table_{out} に変換元のテーブル T i を代入し、children を空集合（ Φ ）で初期化する。parent は、生成対象のノード（ここではルートノード）の親ノードに対応する変数で

ある。ルートノードに親ノードは無いため、ここでは `parent` に `null` が代入される。`edge` は、`parent` と生成対象のノードとを接続するエッジに対応する変数である。ルートノードには該当するエッジは無いため、ここでは `edge` に `null` が代入される。`tableout` は、生成対象のノードに対応するテーブルを格納する変数である。最初は、テーブル `Ti` に対応するルートノードが生成対象とされるため、`tableout` に対してテーブル `Ti` が代入される。

[0032] 続いて、変換部 13 は、`calcCost (tableout, Ep, En)` を実行し、その出力を `cost` に代入する (S212)。`calcCost` は、テーブルがどの程度正例を含み、どの程度負例を含まないかを定量的に評価する評価関数である。評価関数は、評価対象のテーブル、`Ep` が含む全ての正例、`En` が含む全ての負例を入力し、コストを出力する。但し、図2の正例 `Ep` のみが入力されている場合、`En` は空である。評価関数は正例を含めば含むほど高い評価を出力し、負例を含めば含むほど低い評価を出力する関数であれば任意の関数を用いて構わない。上記したように、本実施の形態では、コストは0以上であり、0が最良（最も高い評価）である。

[0033] 続いて、変換部 13 は、`{edge, tableout, cost, parent, children}` を `node` に代入する (S213)。`node` は生成対象のノードである。

[0034] 続いて、変換部 13 は、探索グラフ `G` に対して `node` を追加する (S214)。その結果、探索グラフ `G` にはルートノードが生成される。

[0035] 続くステップ S221～S223 では、評価関数が0になるノードの探索が行われる。

[0036] ステップ S221 において、変換部 13 は、探索グラフ `G` の末端ノードの中から、コストが最良である（コストが0に最も近い）ノードを1つ選択し、当該ノードを `nodeparent` に代入する。

[0037] 続いて、変換部 13 は、`nodeparent` のテーブルを操作対象とし、操作対象のテーブルに対して実行可能な全てのオペレータを実行する (S222)

）。その結果、探索グラフGが変化する。

[0038] 続いて、変換部13は、探索グラフGの中に $cost = 0$ であるノードが有るか否かを判定する（S223）。該当するノードが無い場合（S223でNo）、変換部13は、ステップS221以降を繰り返す。すなわち、変更後の探索グラフGにおいて末端のノードの中でコストが最良であるノードのテーブルが操作対象とされてオペレータの実行が繰り返される。該当するノードが有る場合（S223でYes）。変換部13は、該当するノードに対応するテーブル（ $table_{out}$ ）を返却する（S224）。

[0039] 続いて、ステップS222の詳細について説明する。図8は、ノードに対応するテーブルに対するオペレータの実行処理の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。

[0040] ステップS301において、変換部13は、探索グラフGの中の末端ノード（ $children$ が中のノード）の中で最も $cost$ が小さいノード（コストが最良であるノード）を $parent$ に代入する。つまり、当該ノードがこれから生成されるノードの親ノードとして選択される。最初は、 $node_{parent}$ はテーブル T_i に対応するルートノードである。

[0041] 続いて、変換部13は、 $node_{parent}$ に対応するテーブルに対して実行可能なオペレータの集合 O_e の要素ごとにループ処理L1を実行する。なお、或るテーブルに対して実行可能なオペレータは、オペレータの引数の値も区別される。仮に、 $move_end(0)$ 及び $move_end(1)$ を実行可能であれば、2種類のオペレータが実行可能であることになる。テーブル T_i に対して実行可能なオペレータは、 $move_end(0)$ 及び $split(1, ', ')$ の2種類である。したがって、テーブル T_i が $parent_node$ である場合、これら2種類のオペレータのそれぞれについてループ処理L1が実行される。ループ処理L1では、実行可能なオペレータごとに、 $node_{parent}$ に対して子ノードが生成するためにステップS302～S309を含む。ループ処理L1において処理対象とされているオペレータを、以下「オペレータp」という。

- [0042] ステップS302において、変換部13は、 $node_{parent}$ と生成対象のノードを接続する $edge$ にオペレータ p を代入する。
- [0043] 続いて、変換部13は、 $node_{parent}$ に対応するテーブル($table_{out}$)に対してオペレータ p を実行することで得られるテーブルを $table_{out}$ に代入する(S303)。
- [0044] 続いて、変換部13は、 $calcCost(table_{out}, Ep, En)$ を実行し、その出力を $cost$ に代入する(S304)。つまり、 $table_{out}$ の評価関数の値が $cost$ に代入される。続いて、変換部13は、 $parent$ に $node_{parent}$ を代入し、 $children$ を空集合(Φ)で初期化する(S305)。続いて、変換部13は、 $\{edge, table_{out}, cost, parent, children\}$ を $node$ に代入する(S306)。
- [0045] 続いて、変換部13は、探索グラフGの中に $table_{out}$ と同じテーブルを持つノードが有るか否かを判定する(S307)。該当するノードが無い場合(S307でYes)、変換部13は、探索グラフGに $node$ を追加する(S308)。続いて、変換部13は、 $node_{parent}$ の $children$ に対して $node$ を追加する(S309)。その結果、 $node$ が $node_{parent}$ の子ノードであることが登録される。なお、同じテーブルに対応するノードを重複させないのは、同じテーブルを持つノードの先の探索グラフは全て同じになるため探索に無駄が生じるからである。
- [0046] $node_{parent}$ がテーブル T_i である場合、図8の処理手順の実行により、探索グラフGは、図3に示した通りになる。この場合、テーブル t_2 の $cost$ が0であるため、変換部13は、図6のステップS224においてテーブル t_2 を返却する。出力部14は、テーブル t_2 を出力する。
- [0047] テーブル t_2 は、ユーザが求めるテーブル T_o (図1)とは異なる。具体的にはテーブル t_2 の2列目が空の行も含まれてしまっている。そこで、ユーザは負例によってこれらの行を取り除くようにする。具体的には、ユーザは、図2の負例 En をデータ変換装置10に入力する。負例 En 入力、事例入

力部 12 が図 6 のステップ S 102 において受け付ける。

[0048] 続いて、変換部 13 がデータ変換処理（図 7）を再実行する（S 103）。その結果、図 9 のような探索グラフ G が生成される。図 9 の探索グラフ G においては、テーブル t_5 の cost が 0 である。したがって、変換部 13 は、図 6 のステップ S 224 においてテーブル t_5 を返却する。出力部 14 は、テーブル t_5 を出力する。その結果、ユーザは、所望の変換結果を得ることができる。

[0049] 上述したように、第 1 の実施の形態によれば、データ変換のための作業効率を向上させることができる。具体的には、ユーザは、変換方法やプログラミングの知識がなくても少ない稼働でデータ変換を行うことができる。それにより、データサイエンティスト以外でもデータ分析を行うことが可能となり、日常的な業務にデータ分析によって得られた知見を反映させることができる。

[0050] なお、本実施の形態における変換前のデータは、便宜上、全データ（全レコード）を容易に把握可能な程度の規模であるが、本実施の形態は、大規模なデータに対しての適用も可能であり、その場合にはより顕著に効果を得ることができる。

[0051] 次に、第 2 の実施の形態について説明する。第 2 の実施の形態では第 1 の実施の形態と異なる点について説明する。第 2 の実施の形態において特に言及されない点については、第 1 の実施の形態と同様でもよい。

[0052] 第 1 の実施の形態では、2 回目以降のステップ S 103（図 7）の実行において、毎回ルートノードの生成から行われる。つまり、 n 回目のステップ S 103 においては、 $n - 1$ 回目において探索グラフ G の再構築から実行される。第 2 の実施の形態では、このような非効率を改善した例について説明する。

[0053] 第 2 の実施の形態では、図 6 のステップ S 103 において実行される処理手順が、図 7 から図 10 に置き換わる。

[0054] 図 10 は、第 2 の実施の形態におけるデータ変換処理の処理手順の一例を

説明するためのフローチャートである。図10中、図7と同一ステップには同一ステップ番号を付し、その説明は省略する。

[0055] 図10では、ステップS201が追加されている。ステップS201において、変換部13は、探索グラフGが未生成であるか否かを判定する。ステップS201の判定は、図10の処理手順が1回目の実行であるか2回目以降の実行であるかの判定と等価である。

[0056] 探索グラフGが未生成である場合（図10の処理手順が1回目の実行である場合）（S201でYes）、変換部13は、第1の実施の形態と同じ処理手順を実行する。

[0057] 探索グラフGを生成済みである場合（図10の処理手順が2回目の実行である場合）（S201でNo）、変換部13は、生成済みの探索グラフGの各ノードのコストの再計算処理を実行する（S215）。つまり、新たな正例又は負例の追加に基づいて各ノードのcostの値が再計算される。例えば、図10の処理手順が2回目である場合、探索グラフGは、図3の状態である。この場合、図3の各ノードについて再計算が行われる。

[0058] 続いて、変換部13は、ステップS233以降を実行する。

[0059] 続いて、ステップS215の詳細について説明する。図11は、各ノードのコストの再計算処理の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。

[0060] まず、変換部13は、既に生成されている最新の探索グラフGの全てのnodeごとにループ処理L2を実行する。ループ処理L2は、ステップS401～S403を含む。ループ処理L2において処理対象とされているnodeを、以下「対象node」という。

[0061] ステップS401において、変換部13は、calcCost（対象nodeのtable_{out}, Ep, En）を実行し、その出力をcost_{new}に代入する。なお、Epは最新の全ての正例であり、Enは最新の全ての負例である。したがって、対象nodeに対応するテーブルについて、最新の正例及び最新の負例に基づいてコスト（評価関数の値）が再計算される。続いて

、変換部13は、対象 $node$ の $cost$ に $cost_{new}$ を代入する（S402）。

[0062] 続いて、変換部13は、 $cost_{new}$ が0でないか否かを判定する（S403）。 $cost_{new}$ が0でない場合（S403でYes）、変換部13は、ループ処理L2を継続する。 $cost_{new}$ が0である場合（S403でNo）、変換部13は、ループ処理L2を抜けて、図11の処理手順を終了させる。この場合、コストの再計算により所望のテーブル（全ての正例を含み全ての負例を含まないテーブル）を発見できたため、これ以上コストの再計算を行う必要は無いからである。

[0063] 正例 E_p （図2）のみが与えられている状態において生成された探索グラフ G が図3である状態において、負例 E_n （図 E_n ）が新たに与えられた場合に図11の処理手順が実行されると、探索グラフ G は、例えば、図12のように変更される。

[0064] 図12は、コストの再計算後の探索グラフ G の一例を示す図である。図12においては、テーブル t_2 の $cost$ の値が0から1に変更されている。図10のステップS215に続いて実行されるステップS223では、図12に示される探索グラフ G を用いて変換プログラム（所望の変換結果）の探索が行われる。

[0065] 図11の処理手順を導入する利点は、探索グラフ G における既探索部分の計算がコストの再計算のみであるため、効率的であることである。第1の実施の形態のように探索の度に探索グラフ G を生成しなおす場合と比較して、既存部分に関してテーブルの再取得（図8のS303）、同一テーブルの存在確認（図8S307）等の処理が削減されるため、既探索部分については探索を高速化することができる。

[0066] なお、図10のループ処理L2を、探索グラフ G の全ての $node$ ごとではなく、末端の全ての $node$ ごとのループ処理とすることで、更に高速化が図られてもよい。

[0067] 但し、ユーザが、過去にケアレスミスで誤った正例又は負例を与えていて

、ユーザがその誤りに気付いた場合、当該正例又は当該負例が修正される可能性も有る。その場合、コスト（評価関数）の再計算によって $cost = 0$ となるノードが末端以外発生する可能性が有る。探索グラフ G の全ての $node$ ごとにループ処理 $L2$ を実行することで、そのような場合に対処可能とすることができる。

[0068] 上述したように、第2の実施の形態によれば、既存の探索グラフ G を再利用することで、2回目以降の探索を効率化することができる。それにより効率的な対話を実現し、データ変換を高速で行うことができる。

[0069] 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

[0070]	10	データ変換装置
	11	データ入力部
	12	事例入力部
	13	変換部
	14	出力部
	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
	106	表示装置
	107	入力装置
	B	バス

請求の範囲

- [請求項1] 変換対象のテーブルと、前記テーブルの一部に関する正例の変換結果と負例の変換結果とを入力するように構成されている事例入力部と、
- 、
- 前記変換対象のテーブルを最初の操作対象とし、複数の操作のうち前記操作対象に対して実行可能な操作を実行することで得られるテーブルごとに、前記正例及び前記負例との比較に基づく評価値を計算し、前記評価値が最良であるテーブルを前記操作対象とすることを繰り返すことで、前記正例を含み前記負例を含まないテーブルを生成するように構成されている変換部と、
- 前記正例を含み前記負例を含まないテーブルを出力するように構成されている出力部と、
- を有し、
- 前記変換部は、前記出力部がテーブルを出力した後に、前記事例入力部が新たな正例又は新たな負例を入力した場合に、前記新たな正例をも含み前記新たな負例をも含まないテーブルを生成するように構成されている、
- ことを特徴とするデータ変換装置。
- [請求項2] 前記変換部は、前記出力部がテーブルを出力した後に、前記事例入力部が新たな正例又は新たな負例を入力した場合に、既に生成した一部又は全てのテーブルについて前記評価値を再計算し、前記操作対象の系列の中で末端のテーブルのうち前記評価値が最良であるテーブルを最初の前記操作対象とするように構成されている、
- ことを特徴とする請求項1記載のデータ変換装置。
- [請求項3] 前記変換部は、前記出力部がテーブルを出力した後に、前記事例入力部が新たな正例又は新たな負例を入力した場合に、既に生成した全てのテーブルについて前記評価値を再計算し、前記操作対象の系列の中で末端のテーブルのうち前記評価値が最良であるテーブルを最初の

前記操作対象とするように構成されている、
ことを特徴とする請求項2記載のデータ変換装置。

[請求項4] 前記変換部は、前記出力部がテーブルを出力した後に、前記事例入力部が新たな正例又は新たな負例を入力した場合に、前記操作対象の系列の中で末端のテーブルについてのみ前記評価値を再計算し、前記末端のテーブルのうち前記評価値が最良であるテーブルを最初の前記操作対象とするように構成されている、
ことを特徴とする請求項2記載のデータ変換装置。

[請求項5] 変換対象のテーブルと、前記テーブルの一部に関する正例の変換結果と負例の変換結果とを入力する事例入力手順と、

前記変換対象のテーブルを最初の操作対象とし、複数の操作のうち前記操作対象に対して実行可能な操作を実行することで得られるテーブルごとに、前記正例及び前記負例との比較に基づく評価値を計算し、前記評価値が最良であるテーブルを前記操作対象とすることを繰り返すことで、前記正例を含み前記負例を含まないテーブルを生成する変換手順と、

前記正例を含み前記負例を含まないテーブルを出力する出力手順と、
、
をコンピュータが実行し、

前記変換手順は、前記出力手順がテーブルを出力した後に、前記事例入力手順が新たな正例又は新たな負例を入力した場合に、前記新たな正例をも含み前記新たな負例をも含まないテーブルを生成する、
ことを特徴とするデータ変換方法。

[請求項6] 変換対象のテーブルと、前記テーブルの一部に関する正例の変換結果と負例の変換結果とを入力する事例入力手順と、

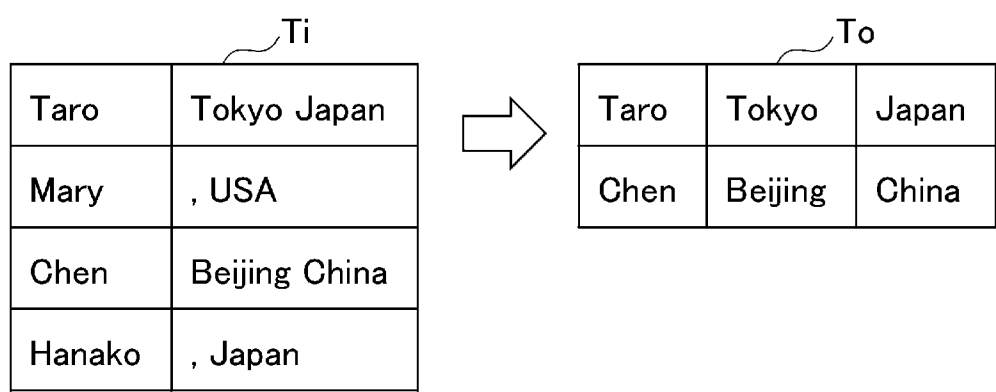
前記変換対象のテーブルを最初の操作対象とし、複数の操作のうち前記操作対象に対して実行可能な操作を実行することで得られるテーブルごとに、前記正例及び前記負例との比較に基づく評価値を計算し

、前記評価値が最良であるテーブルを前記操作対象とすることを繰り返すことで、前記正例を含み前記負例を含まないテーブルを生成する変換手順と、

前記正例を含み前記負例を含まないテーブルを出力する出力手順と、
、
をコンピュータに実行させ、

前記変換手順は、前記出力手順がテーブルを出力した後に、前記事例入力手順が新たな正例又は新たな負例を入力した場合に、前記新たな正例をも含み前記新たな負例をも含まないテーブルを生成する、ことを特徴とするプログラム。

[図1]



[図2]

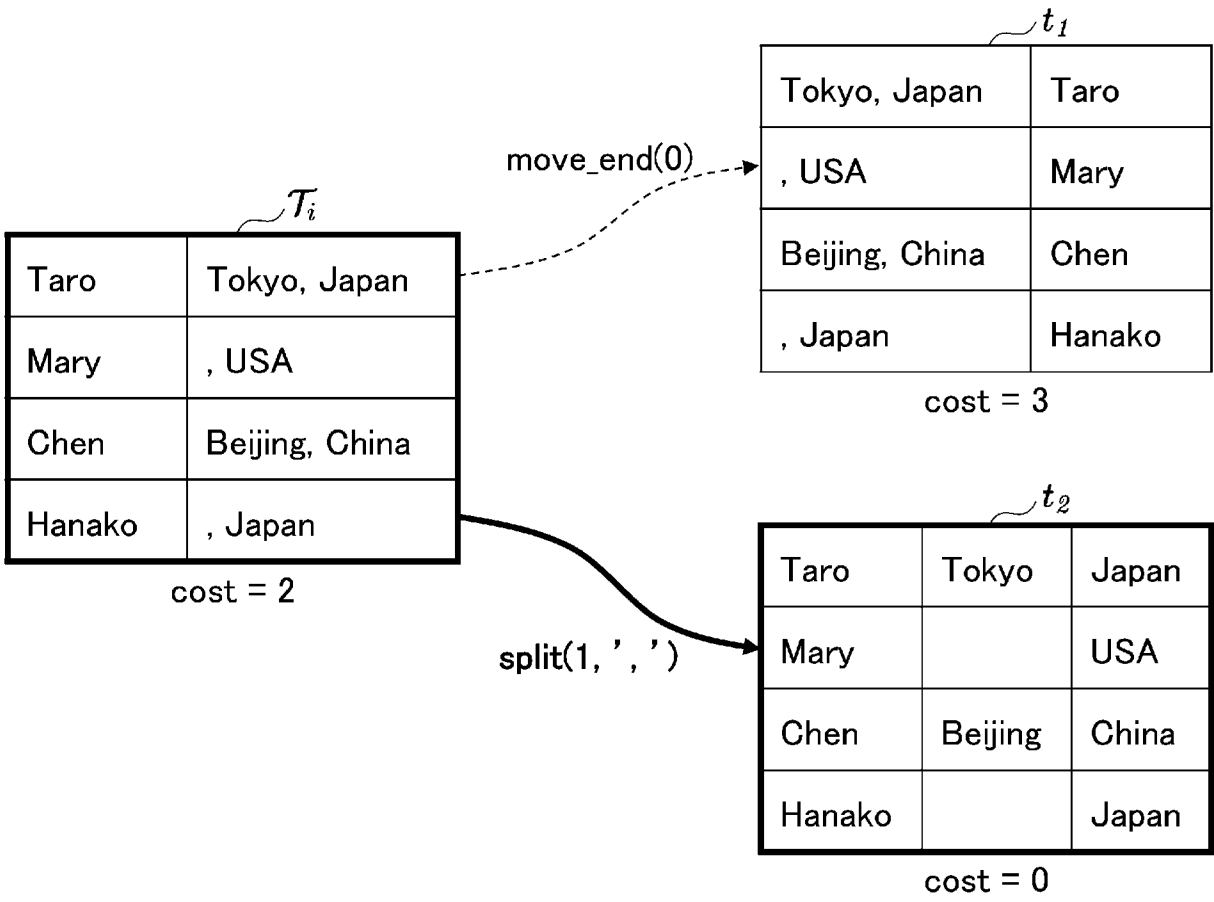
正例Ep

Taro	Tokyo	Japan
------	-------	-------

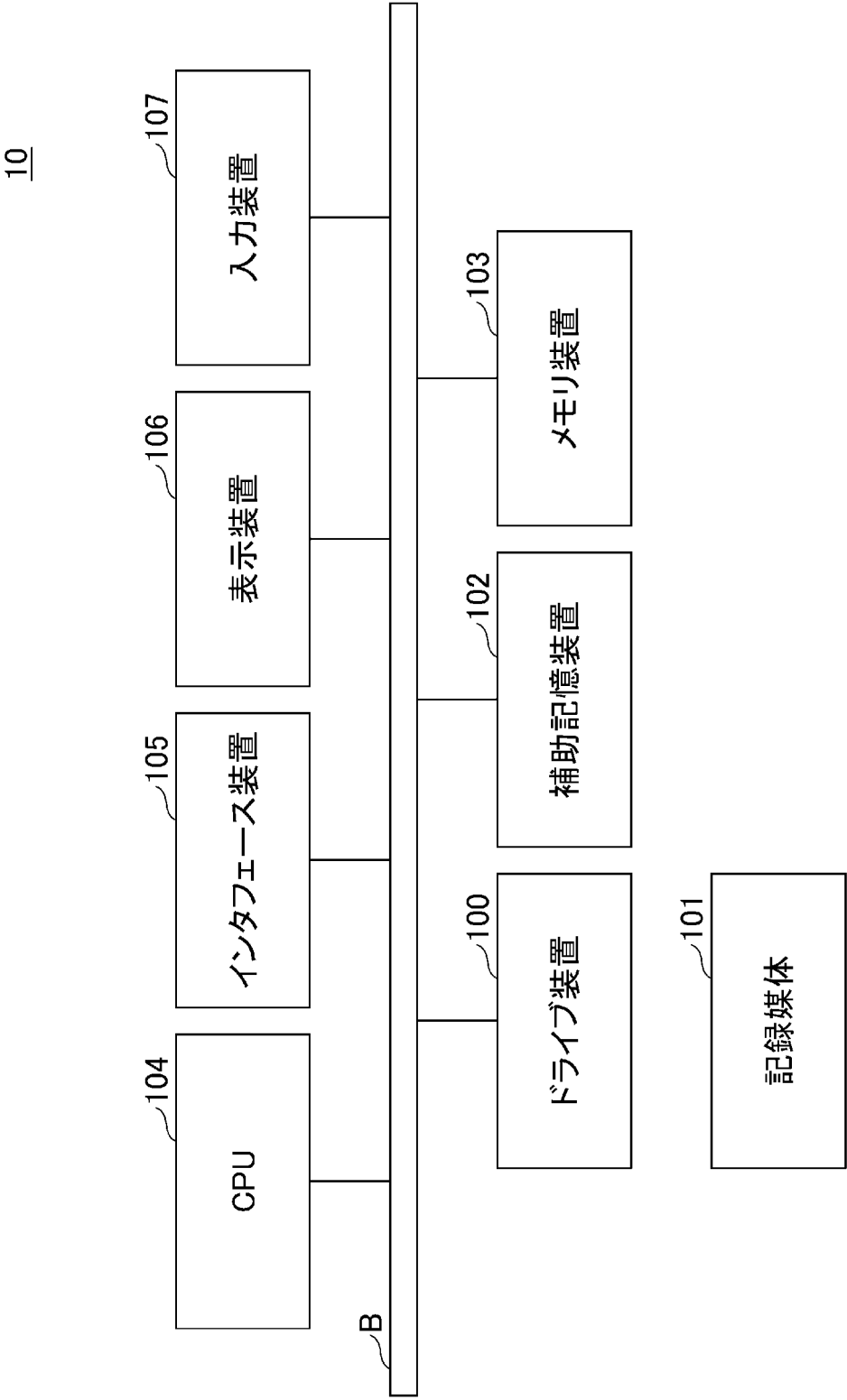
負例En

Mary		USA
------	--	-----

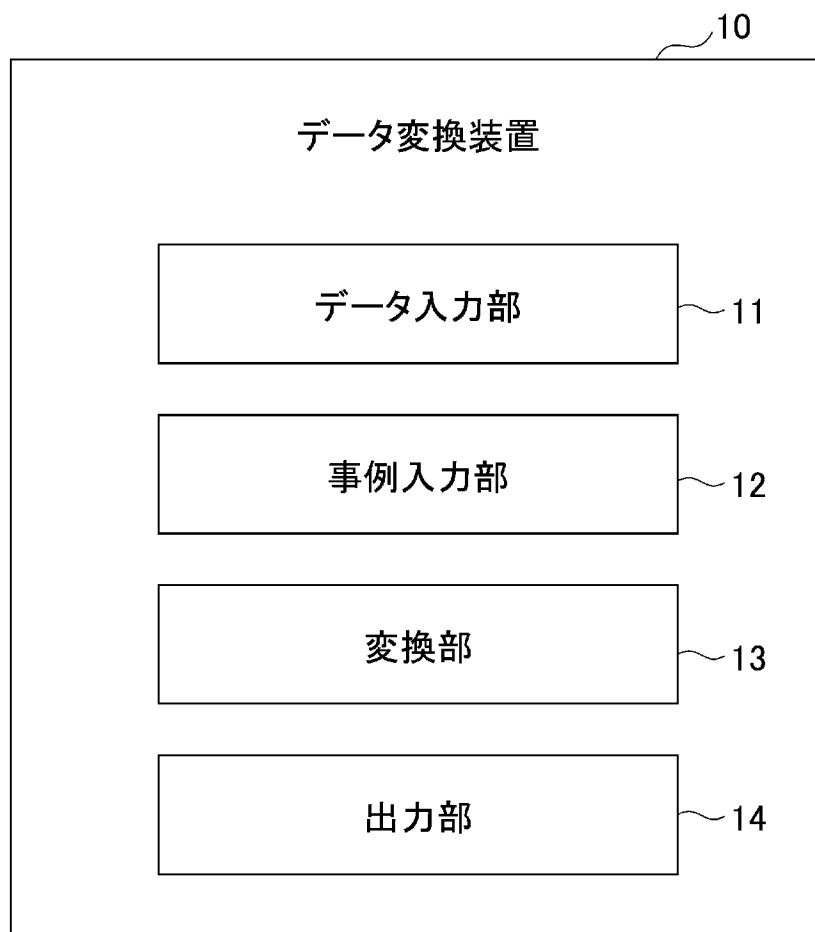
[図3]



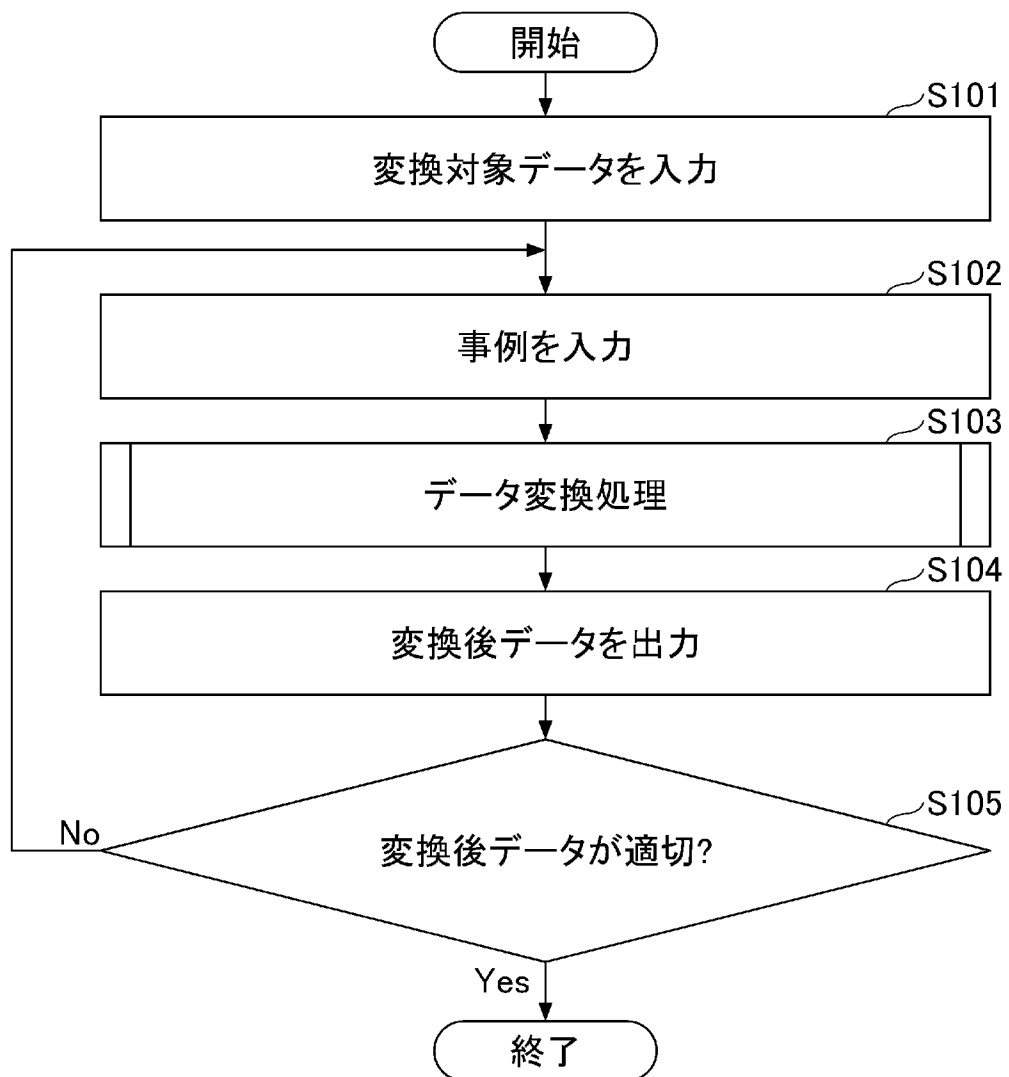
[図4]



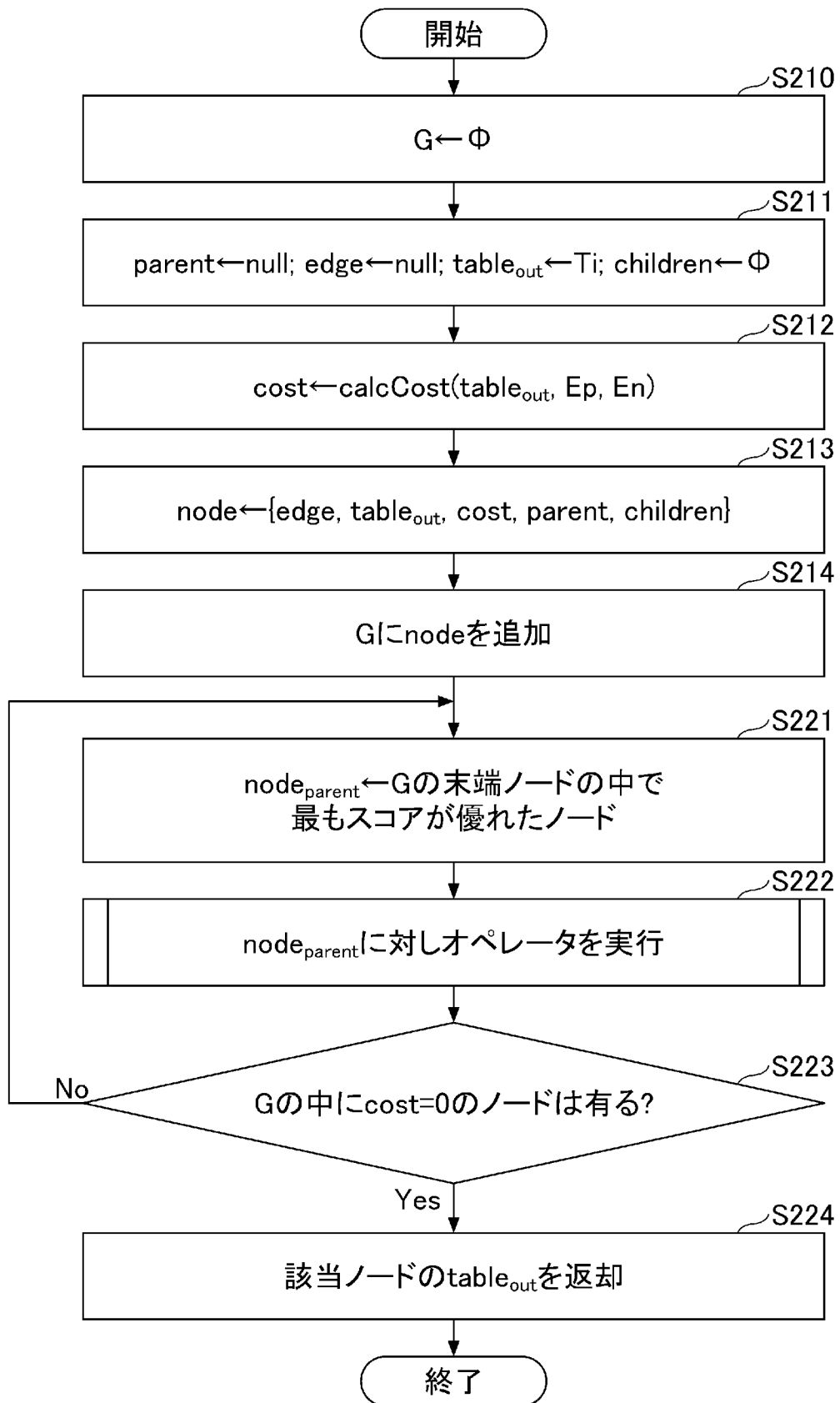
[図5]



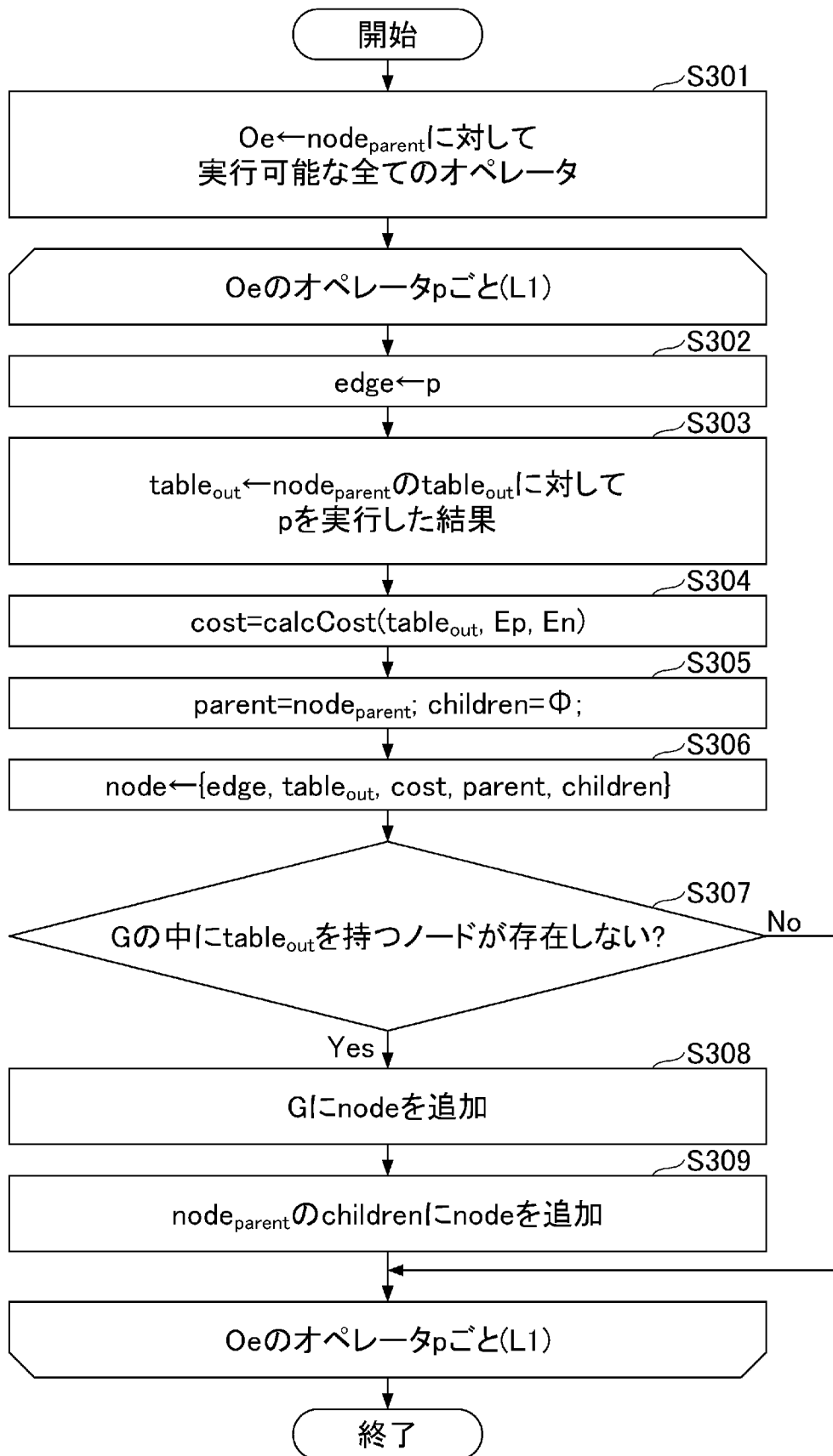
[図6]



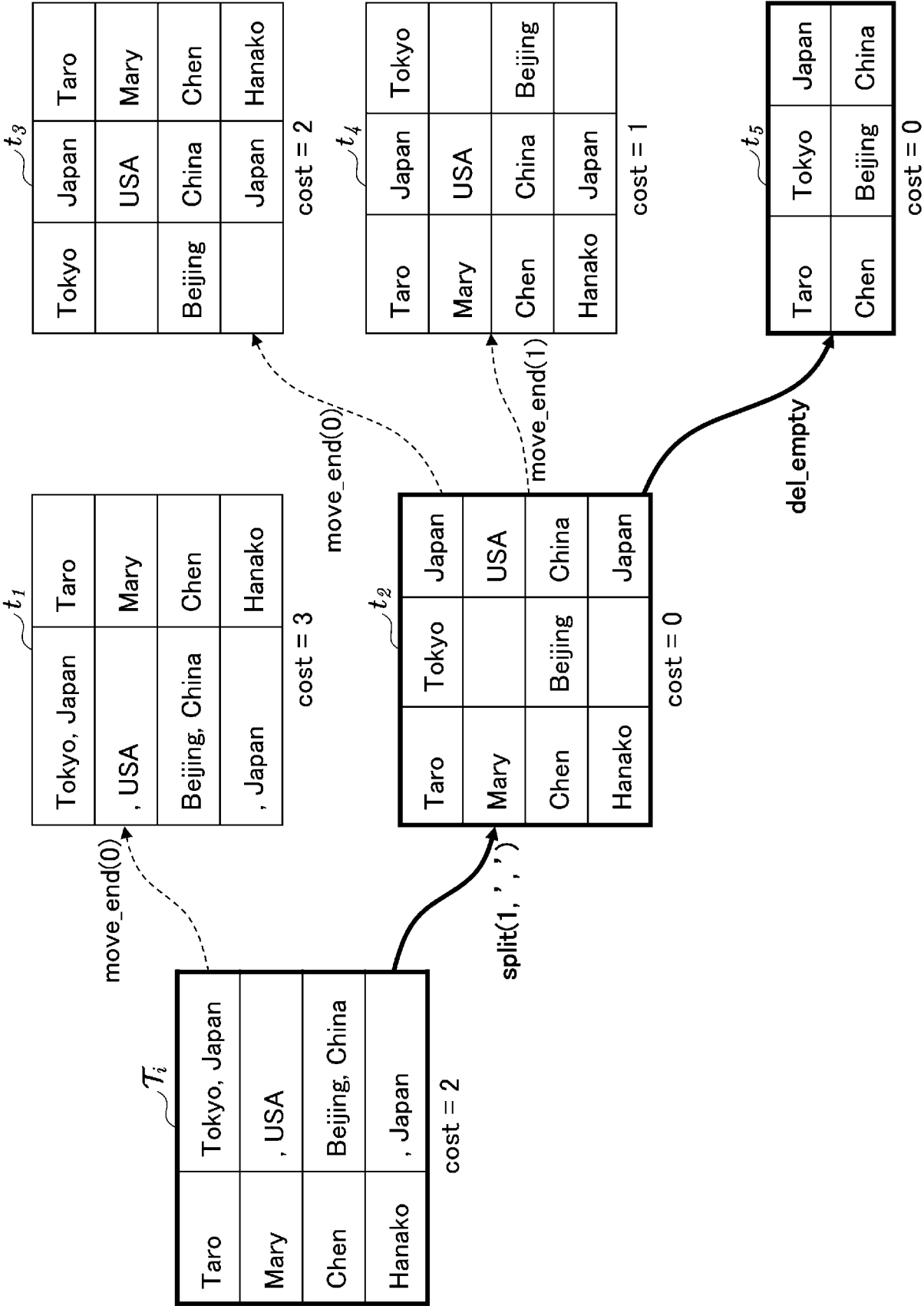
[図7]



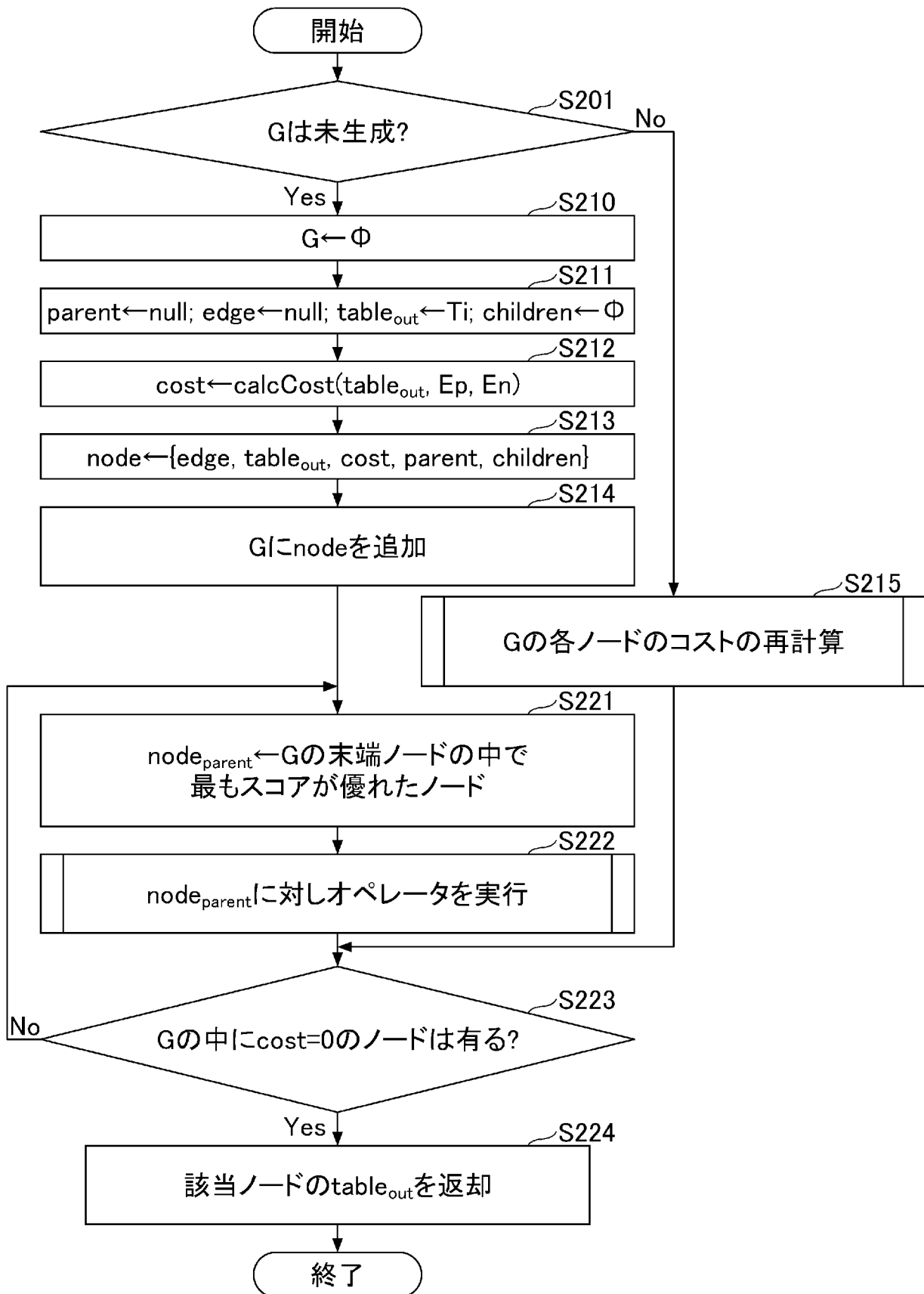
[図8]



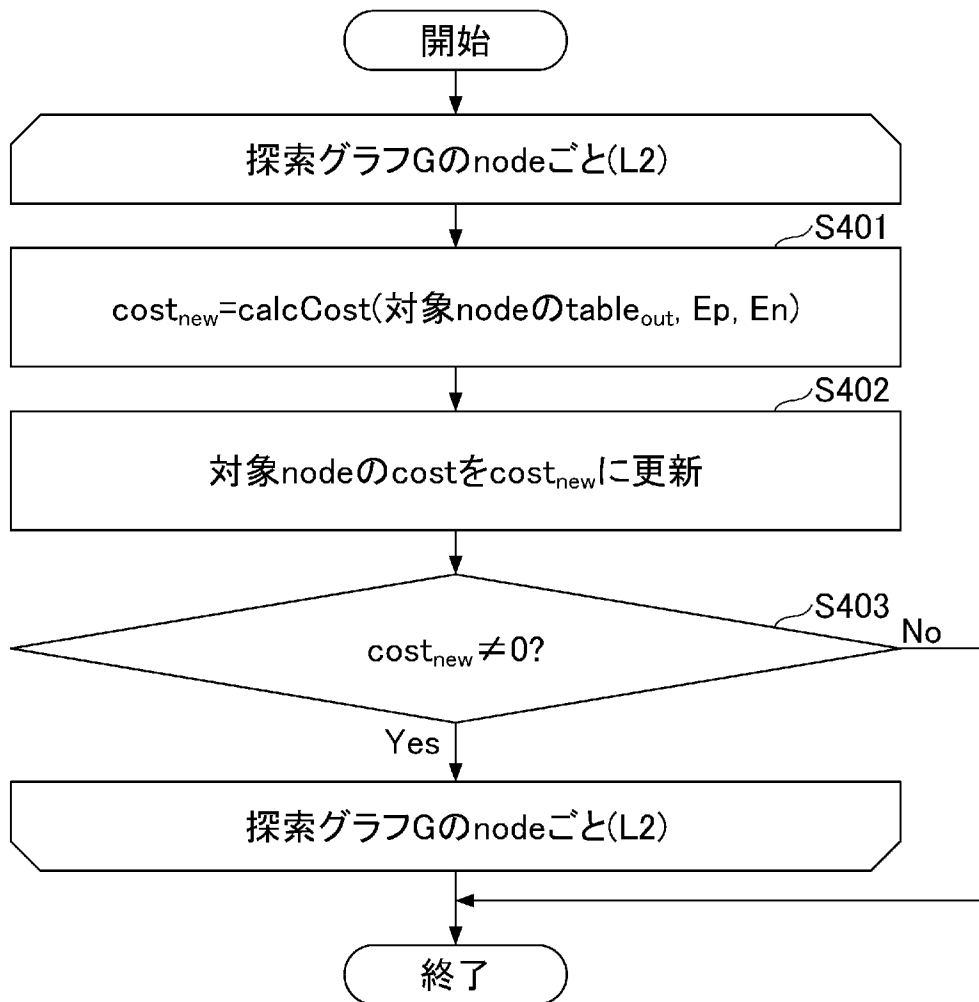
[図9]



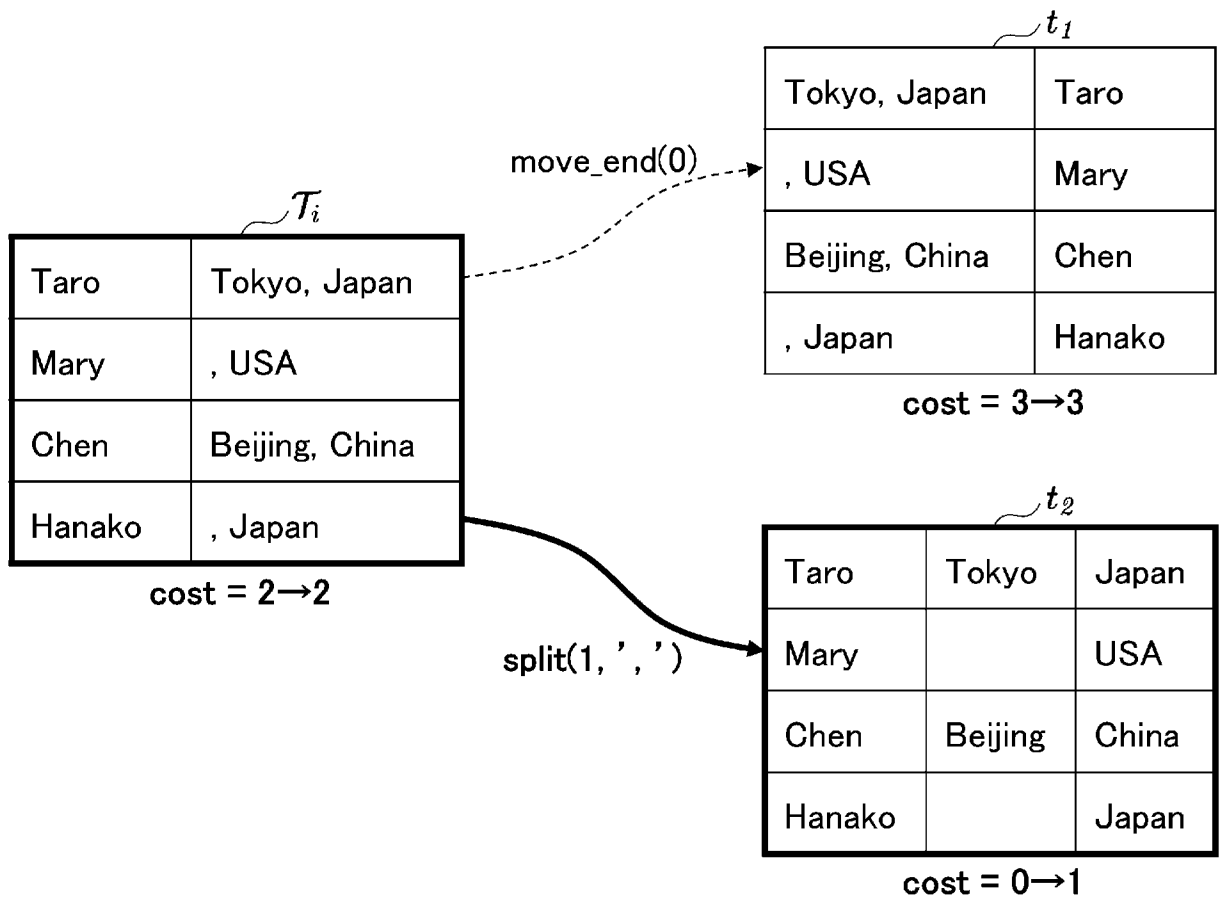
[図10]



[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/020846

A. CLASSIFICATION OF SUBJECT MATTER**G06F 16/21**(2019.01)i; **G06F 40/16**(2020.01)i

FI: G06F16/21; G06F40/16

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)

G06F16/00-16/29; G06F40/00-40/58

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 1971-2023
 Registered utility model specifications of Japan 1996-2023
 Published registered utility model applications of Japan 1994-2023

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	US 2015/0095312 A1 (GULWANI, Sumit) 02 April 2015 (2015-04-02)	1-6
A	US 2011/0302553 A1 (GULWANI, Sumit) 08 December 2011 (2011-12-08)	1-6
A	US 2013/0144902 A1 (GULWANI, Sumit) 06 June 2013 (2013-06-06)	1-6

☐ 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

21 July 2023

Date of mailing of the international search report

01 August 2023

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.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2023/020846

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2015/0095312	A1	02 April 2015	WO	2015/050909	A1	
US	2011/0302553	A1	08 December 2011	WO	2011/153006	A2	
				EP	2577460	A2	
				CN	102918500	A	
				KR	10-2013-0111211	A	
US	2013/0144902	A1	06 June 2013	(Family: none)			

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/21(2019.01)i; G06F 40/16(2020.01)i FI: G06F16/21; G06F40/16										
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）） G06F16/00-16/29; G06F40/00-40/58										
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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2023年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2023年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2023年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2023年	日本国実用新案登録公報	1996 - 2023年	日本国登録実用新案公報	1994 - 2023年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2023年									
日本国実用新案登録公報	1996 - 2023年									
日本国登録実用新案公報	1994 - 2023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	US 2015/0095312 A1 (GULWANI, Sumit) 02.04.2015 (2015 - 04 - 02)	1-6								
A	US 2011/0302553 A1 (GULWANI, Sumit) 08.12.2011 (2011 - 12 - 08)	1-6								
A	US 2013/0144902 A1 (GULWANI, Sumit) 06.06.2013 (2013 - 06 - 06)	1-6								
☐ 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。										
<table border="0"> <tr> <td> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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 </td> <td> “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 </td> </tr> </table>			* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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						
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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 21.07.2023	国際調査報告の発送日 01.08.2023									
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） 甲斐 哲雄 5N 9750 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/JP2023/020846

引用文献			公表日	パテントファミリー文献			公表日
US	2015/0095312	A1	02.04.2015	WO	2015/050909	A1	
US	2011/0302553	A1	08.12.2011	WO	2011/153006	A2	
				EP	2577460	A2	
				CN	102918500	A	
				KR	10-2013-0111211	A	
US	2013/0144902	A1	06.06.2013	(ファミリーなし)			