

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6154491号  
(P6154491)

(45) 発行日 平成29年6月28日(2017.6.28)

(24) 登録日 平成29年6月9日(2017.6.9)

(51) Int. Cl.	F I
<b>G06F 17/18 (2006.01)</b>	G06F 17/18 Z
<b>G06F 17/16 (2006.01)</b>	G06F 17/16 Q

請求項の数 10 (全 23 頁)

(21) 出願番号	特願2015-559717 (P2015-559717)	(73) 特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(86) (22) 出願日	平成26年2月3日(2014.2.3)	(74) 代理人	110001678 特許業務法人藤央特許事務所
(86) 国際出願番号	PCT/JP2014/052441	(72) 発明者	宮本 篤志 東京都千代田区丸の内一丁目6番6号 株 式会社日立製作所内
(87) 国際公開番号	W02015/114830	(72) 発明者	工藤 泰幸 東京都千代田区丸の内一丁目6番6号 株 式会社日立製作所内
(87) 国際公開日	平成27年8月6日(2015.8.6)	(72) 発明者	宮越 純一 東京都千代田区丸の内一丁目6番6号 株 式会社日立製作所内
審査請求日	平成28年6月2日(2016.6.2)		

最終頁に続く

(54) 【発明の名称】 計算機及びグラフデータ生成方法

(57) 【特許請求の範囲】

【請求項1】

プロセッサ、及び前記プロセッサに接続されるメモリを備え、複数の指標の間の相関関係を示す値を要素とする相関行列データを用いた処理を実行する計算機であって、

前記計算機は、記憶装置から取得される前記相関行列データから、一つの指標に対応する頂点、相関関係のある二つの前記頂点を接続するエッジ、及び前記要素の値であるエッジの重みから構成されるグラフデータを生成するグラフ処理部を備え、

前記グラフ処理部は、

前記相関行列データを用いた処理に対する制約条件に基づいて、圧縮された前記グラフデータを生成するための制御因子を算出する制御因子算出部と、

前記制御因子を用いて、前記相関行列データから前記グラフデータを生成するグラフデータ生成部と、を含み、

前記制御因子算出部は、

前記相関行列データを用いた処理の完了時間である目標処理時間を前記制約条件として取得し、

前記目標処理時間内に処理を完了するために、前記グラフデータに含めることが可能な最大エッジ数を算出し、

前記最大エッジ数に基づいて、前記相関行列データから前記グラフデータに変換する要素を抽出するための閾値を算出し、

前記閾値を前記制御因子として前記グラフデータ生成部に出力し、

10

20

前記グラフデータ生成部は、  
 前記閾値に基づいて、前記相関行列データから一つ以上の要素を抽出し、  
 抽出された前記一つ以上の要素を識別するための前記指標の組合せ及び当該要素の値から、前記頂点、前記エッジ及び前記エッジの重みの情報を生成することによって前記グラフデータを生成することを特徴とする計算機。

【請求項 2】

請求項 1 に記載の計算機であって、  
 前記グラフデータ生成部は、  
 前記相関行列データの要素の値の大きさが前記閾値より小さい前記要素の値を零に設定し、  
 前記相関行列データから前記要素の値が零でない前記要素を抽出することを特徴とする計算機。

10

【請求項 3】

請求項 1 に記載の計算機であって、  
 前記グラフデータ生成部は、  
 前記相関行列データから、前記要素の値の大きさが前記閾値より大きい前記要素を抽出することを特徴とする計算機。

【請求項 4】

請求項 2 又は請求項 3 に記載の計算機であって、  
 前記グラフ処理部は、前記相関行列データの要素の値を解析することによって前記相関行列データの要素の値の度数分布を生成するエッジ情報量算出部を含み、  
 前記制御因子算出部は、  
 前記エッジの数と、前記相関行列データを用いた処理の完了時間との関係を示す関数を算出し、  
 前記関数及び前記目標処理時間に基づいて前記最大エッジ数を算出し、  
 前記度数分布及び前期最大エッジ数に基づいて前記閾値を算出することを特徴とする計算機。

20

【請求項 5】

請求項 4 に記載の計算機であって、  
 前記制御因子算出部は、  
 さらに、前記メモリの使用制限量を前記制約条件として取得し、  
 前記メモリの使用制限量に基づいて、前記グラフデータの大きさが前記メモリの使用制限量より小さくなるように前記エッジの重みの表現ビット数を決定し、  
 前記閾値及び前記エッジの重みの表現ビット数を前記制御因子として出力し、  
 前記グラフデータ生成部は、前記エッジの重みを、前記決定された表現ビット数に丸めることを特徴とする計算機。

30

【請求項 6】

プロセッサ、及び前記プロセッサに接続されるメモリを備え、複数の指標の間の相関関係を示す値を要素とする相関行列データを用いた処理を実行する計算機におけるグラフデータ生成方法であって、  
 前記計算機は、記憶装置から取得される前記相関行列データから、一つの指標に対応する頂点、相関関係のある二つの前記頂点を接続するエッジ、及び前記要素の値であるエッジの重みから構成されるグラフデータを生成するグラフ処理部を備え、  
 前記グラフ処理部は、  
 前記相関行列データを用いた処理に対する制約条件に基づいて、圧縮された前記グラフデータを生成するための制御因子を算出する制御因子算出部と、  
 前記制御因子を用いて、前記相関行列データから前記グラフデータを生成するグラフデータ生成部と、を含み、  
 前記グラフデータ生成方法は、  
 前記制御因子算出部が、前記相関行列データを用いた処理の完了時間である目標処理時

40

50

間を前記制約条件として取得する第1のステップと、

前記制御因子算出部が、前記目標処理時間内に処理を完了するために、前記グラフデータに含めることが可能な最大エッジ数を算出する第2のステップと、

前記制御因子算出部が、前記最大エッジ数に基づいて、前記相関行列データから前記グラフデータに変換する要素を抽出するための閾値を算出する第3のステップと、

前記制御因子算出部が、前記閾値を前記制御因子として前記グラフデータ生成部に出力する第4のステップと、

前記グラフデータ生成部が、前記閾値に基づいて、前記相関行列データから一つ以上の要素を抽出する第5のステップと、

前記グラフデータ生成部が、抽出された前記一つ以上の要素を識別するための前記指標の組合せ及び当該要素の値から、前記頂点、前記エッジ及び前記エッジの重みの情報を生成することによって前記グラフデータを生成する第6のステップと、を含むことを特徴とするグラフデータ生成方法。

10

【請求項7】

請求項6に記載のグラフデータ生成方法であって、

前記第5のステップは、

前記相関行列データの要素の値の大きさが前記閾値より小さい前記要素の値を零に設定するステップと、

前記相関行列データから前記要素の値が零でない前記要素を抽出するステップと、を含むことを特徴とするグラフデータ生成方法。

20

【請求項8】

請求項6に記載のグラフデータ生成方法であって、

前記第5のステップは、前記相関行列データから、前記要素の値の大きさが前記閾値より大きい前記要素を抽出するステップを含むことを特徴とするグラフデータ生成方法。

【請求項9】

請求項7又は請求項8に記載のグラフデータ生成方法であって、

前記グラフ処理部は、前記相関行列データの要素の値を解析することによって前記相関行列データの要素の値の度数分布を生成するエッジ情報量算出部を含み、

前記第1のステップは、前記エッジの数と、前記相関行列データを用いた処理の完了時間との関係を示す関数を算出するステップを含み、

30

前記第2のステップは、前記関数及び前記目標処理時間に基づいて前記最大エッジ数を算出するステップを含み、

前記第3のステップは、前記度数分布及び前期最大エッジ数に基づいて前記閾値を算出するステップを含むことを特徴とするグラフデータ生成方法。

【請求項10】

請求項9に記載のグラフデータ生成方法であって、

前記第1のステップでは、さらに、前記メモリの使用制限量を前記制約条件として取得し、

前記第2のステップは、前記メモリの使用制限量に基づいて、前記グラフデータの大きさが前記メモリの使用制限量より小さくなるように前記エッジの重みの表現ビット数を決定するステップを含み、

40

前記第4のステップでは、前記閾値及び前記エッジの重みの表現ビット数を前記制御因子として出力し、

前記第6のステップは、前記エッジの重みを、前記決定された表現ビット数に丸めるステップを含むことを特徴とするグラフデータ生成方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、グラフデータを用いたビックデータ解析における計算機及びグラフデータ生成方法に関する。

50

## 【背景技術】

## 【0002】

Webやセンサ等から得られた大量のデータ(ビッグデータ)を使用して、有用な知見(情報)を抽出するビッグデータ解析が注目されている。ビッグデータ解析では、統計学、パターン認識、及び人工知能等のデータ解析の技法を大量のデータに網羅的に適用することによって、データの中に潜む項目間の相関関係及びパターンを、知識として抽出する。データの中に隠れた潜在的な情報を「採掘(mining)」することから、ビッグデータ解析はデータマイニングとも呼ばれる。ビッグデータ解析の技法としては、例えば、統計学における相関分析、回帰分析、及び主成分分析、並びに、パターン認識、人工知能における機械学習、及びクラスタリング等がある。

10

## 【0003】

ビッグデータ解析において有用な知識を得るためには、膨大なデータの解析する必要がある。しかし、データ量の増加及びデータ解析方法の高度化に伴って、処理時間及びメモリ使用量の増加等がハードウェアリソースに対して過度な負担がかかることが課題となっている。特に、社会インフラ分野では、限られたハードウェアリソースを用いて、限られた時間内に効率的に結果を出力することが求められる。

## 【0004】

例えば、統計的なデータ解析手法として基本的な相関分析及び主成分分析では、ビッグデータから指標(特徴量、項目)を生成し、指標間の相関関係を導出する。このとき、指標の数が $m$ である相関関係は $m$ 列 $m$ 行の相関行列として与えられ、相関分析及び主成分分析は、相関行列の演算によって実行される。しかし、行列演算は全ての要素について演算処理を実行するために、全ての要素のデータを蓄積しなければならない。そのため、ビッグデータを扱うシステムでは、計算量、及びメモリ使用量の観点から非常に効率が悪くなる。その結果、大量の指標から構成されるビッグデータ(相関行列)の蓄積及び演算処理は、ハードウェアリソースに対して大きな負担となる。

20

## 【0005】

ビッグデータを圧縮して処理を効率化する方法としては、米国特許出願公開第2001/0011958号明細書(特許文献1)に記載の技術がある。特許文献1には、データの蓄積と通信コストの低減を目的に、ビッグデータを多変量のデータ解析手法を用いて変換し、圧縮・再構成する技術が開示されている。特許文献1に開示されている方法は、サンプル数 $n$ 、指標数 $m$ として、 $n$ 列 $m$ 項目のオリジナルデータから $m$ 列 $m$ 行の相関行列を得るステップと、相関行列の固有値と固有ベクトルを求めるステップと、固有値と固有ベクトルから因子負荷量の行列を求めるステップと、1列 $p$ 行のランダム行列を生成するステップと、ランダム行列と因子負荷量の行列の乗算によって1列 $m$ 行の中間データ行列を得るステップと、中間データ行列をスケールすることで1列 $m$ 行の再構成されたデータ行列を得るステップと、を有する。データの再構成を可能とすることによって、通信及びデータの蓄積のコストを削減することができることが記載されている。

30

## 【先行技術文献】

## 【特許文献】

## 【0006】

【特許文献1】米国特許出願公開第2001/0011958号明細書

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0007】

特許文献1に記載の方法は、データの蓄積及び通信のコストを低減するために、オリジナルデータのサンプル数 $n$ を圧縮することを主な課題としており、分析処理時におけるハードウェアリソースの制約までは十分に考慮されていない。また、特許文献1記載の方法では、相関分析又は主成分分析を行う場合、圧縮されたデータ列を再構築し、元のフォーマットに変換してから相関行列の演算を行い、その後分析処理を実行する必要がある。そのため、特許文献1の方法では、指標数 $m$ はサンプル数 $n$ に対して十分に小さいことが

50

前提とされている。

【0008】

指標数 $m$ の増大に伴って、 $m$ 列 $m$ 行の相関行列がメモリに格納できないほど大きい場合、相関分析又は主成分分析などのデータ解析を行えなくなるという課題がある。社会インフラシステムの分析などにおいては、説明指標が100万規模になることも想定されるため、指標数の増加に伴う分析処理の効率化が必要不可欠となる。

【0009】

本発明は、上述のような課題を解決するためになされたもので、大量の指標から構成される相関行列の分析処理において、データ量を圧縮することによって処理量を削減し、効率化することを目的とする。

【課題を解決するための手段】

【0010】

本願において開示される発明の代表的な一例を示せば以下の通りである。すなわち、プロセッサ、及び前記プロセッサに接続されるメモリを備え、複数の指標の間の相関関係を示す値を要素とする相関行列データを用いた処理を実行する計算機であって、前記計算機は、記憶装置から取得される前記相関行列データから、一つの指標に対応する頂点、相関関係のある二つの前記頂点を接続するエッジ、及び前記要素の値であるエッジの重みから構成されるグラフデータを生成するグラフ処理部を備え、前記グラフ処理部は、前記相関行列データを用いた処理に対する制約条件に基づいて、圧縮された前記グラフデータを生成するための制御因子を算出する制御因子算出部と、前記制御因子を用いて、前記相関行列データから前記グラフデータを生成するグラフデータ生成部と、を含み、前記制御因子算出部は、前記相関行列データを用いた処理の完了時間である目標処理時間を前記制約条件として取得し、前記目標処理時間内に処理を完了するために、前記グラフデータに含めることが可能な最大エッジ数を算出し、前記最大エッジ数に基づいて、前記相関行列データから前記グラフデータに変換する要素を抽出するための閾値を算出し、前記閾値を前記制御因子として前記グラフデータ生成部に出力し、前記グラフデータ生成部は、前記閾値に基づいて、前記相関行列データから一つ以上の要素を抽出し、抽出された前記一つ以上の要素を識別するための前記指標の組合せ及び当該要素の値から、前記頂点、前記エッジ及び前記エッジの重みの情報を生成することによって前記グラフデータを生成することを特徴とする。

【発明の効果】

【0011】

本発明によれば、制約条件に従って、大量の指標から構成される相関行列データを圧縮されたグラフデータに変換することができる。これによって、データ量を削減し、相関分析又は主成分分析等の高速なグラフ処理が可能となる。

【0012】

上記した以外の課題、構成及び効果は、以下の実施例の説明により明らかにされる。

【図面の簡単な説明】

【0013】

【図1】本発明の実施例1のグラフ処理装置の構成例を示すブロック図である。

【図2】本発明の実施例1のグラフ処理装置が適用されるシステム構成の一例を示すブロック図である。

【図3】本発明の実施例1における業務データの一例を示す説明図である。

【図4】本発明の実施例1における相関行列データの一例を示す説明図である。

【図5】本発明の実施例1のグラフ処理装置が実行する処理の概要を説明するフローチャートである。

【図6】本発明の実施例1のエッジ情報量算出処理の一例を説明するフローチャートである。

【図7A】本発明の実施例1の相関値の頻度分布表の一例を示す説明図である。

【図7B】本発明の実施例1のエッジ情報量の一例を示す説明図である。

10

20

30

40

50

【図 8】本発明の実施例 1 の制御因子算出処理の一例を説明するフローチャートである。

【図 9】本発明の実施例 1 の推定処理時間関数  $f(E)$  の一例を示す説明図である。

【図 10】本発明の実施例 1 の制御因子の決定時に用いられる推定用エッジ情報量の一例を示す説明図である。

【図 11】本発明の実施例 1 のグラフデータ生成処理の一例を説明するフローチャートである。

【図 12 A】本発明の実施例 1 のグラフデータ生成処理に用いられる頂点リストの一例を示す説明図である。

【図 12 B】本発明の実施例 1 のグラフデータ生成処理に用いられるエッジリストの一例を示す説明図である。

10

【図 13】本発明の実施例 1 のグラフデータ生成処理における制御因子を用いた相関値の切り捨ての概念を示す説明図である。

【図 14 A】本発明の実施例 1 のグラフデータ生成処理の実行後の頂点リスト及びエッジリストを示す説明図である。

【図 14 B】本発明の実施例 1 のグラフデータ生成処理の実行後の頂点リスト及びエッジリストを示す説明図である。

【図 15】本発明の実施例 1 のグラフデータに基づいて表示されるグラフの一例を示す説明図である。

【図 16】本発明の実施例 2 のグラフ処理装置の構成例を示すブロック図である。

【図 17】本発明の実施例 2 の制御因子算出処理の一例を説明するフローチャートである

20

【図 18 A】本発明の実施例 2 の推定メモリ使用量関数  $g(E, B)$  の一例を示す説明図である。

【図 18 B】本発明の実施例 2 の推定メモリ使用量関数  $g(E, B)$  の一例を示す説明図である。

【図 19】本発明の実施例 2 の相関値の表現ビット数の丸めの一例を示す説明図である。

【発明を実施するための形態】

【0014】

以下、添付図面を参照して本発明の実施例について説明する。添付図面では、機能的に同じ要素は同じ番号で表示されている。なお、添付図面は本発明の原理に則った具体的な実施例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。

30

【0015】

まず、本発明の概要について説明する。

【0016】

業務データに対する相関分析等の分析処理の実行時には、業務データから指標（特徴量、項目等）間の相関関係を示す相関行列データが生成される。指標の数が  $m$  個の場合、相関行列データは  $m$  行  $m$  列の行列データとなる。相関行列データは行列の要素を識別する指標の組合せ、及び要素の値から構成されるデータである。

【0017】

40

ビックデータ解析では指標の数が多いため、相関行列データのサイズも大きい。そのため、メモリに相関行列データを格納することができない。そのため、業務データの解析処理の実行時には、相関行列データを取得するためにストレージ装置等に頻繁にアクセスする必要がある。したがって、ストレージ装置へのアクセスに伴う処理遅延が発生する。

【0018】

また、 $m$  行  $m$  列の相関行列データは  $(m \times m)$  個の要素を持ち、解析処理では、全ての要素のデータを処理する必要がある。指標間の相関がないことを示す値「0」の場合であっても、「0」という値を保持する必要がある。そのため、指標の数が増大すると、処理コスト及びデータ量が増大する。

【0019】

50

### (1) グラフデータへの変換

前述した課題を解決するために、本発明のグラフ処理装置100(図1参照)は、相関行列データをグラフデータに変換する。ここで、グラフデータは、指標を表す頂点、相関のある二つの頂点を接続するエッジ、及び要素の値を表すエッジの重みから構成されるデータ構造のデータであり、頂点間の接続関係をグラフとして把握することができる。エッジの重みが当該エッジで接続される二つの指標の間の相関関係の強さを表す。

#### 【0020】

相関関係が存在しない頂点との間にはエッジが存在しないため、グラフデータでは、相関関係がないことを示すデータを保持する必要がない。また、いずれの頂点とも接続されない場合はデータとして保持する必要がない。一方、相関行列データでは、二つの指標の間に相関関係がない場合であっても「0」を値とする要素としてデータを保持する必要がある。そのため、グラフデータは、相関行列データよりデータ量が少ない。

10

#### 【0021】

したがって、相関行列データをグラフデータに変換することによって、データ量を削減することができる。本発明では、グラフ処理装置100は、単に、相関行列データをグラフデータに変換するのではなく、グラフ処理の制約条件に基づいて圧縮されたグラフデータに変換する点に特徴がある。具体的には、以下の2つの処理を含むことに特徴がある。

#### 【0022】

### (2) グラフデータに含めるエッジ数の調整

相関行列データをそのままグラフデータに変換しても、十分にデータ量を削減することができない可能性がある。そのため、本発明のグラフ処理装置100(図1参照)は、解析処理の処理完了時間である目標処理時間に応じて、グラフデータに含まれるエッジ数を調整する。

20

#### 【0023】

具体的には、グラフ処理装置100は、目標処理時間に基づいて、相関値を切り捨てるための閾値を決定する。さらに、グラフ処理装置100は、各要素の値の大きさ(絶対値)が閾値以下の要素の値を「0」に設定し、その上で、グラフデータに変換する。前述したように、「0」は二つの指標間に相関関係がないことを示し、また、相関関係がない場合にはエッジも存在しない。そのため、グラフデータに含まれるエッジ数を削減することができる。

30

#### 【0024】

### (3) エッジの重みの表現ビット数の丸め

本発明のグラフ処理装置100は、メモリ容量に応じて、エッジの重みの表現ビット数を丸める。これによって、グラフデータを、メモリに格納可能なデータサイズに更に圧縮する。

#### 【0025】

以上のような処理を実行することによって、処理に必要なデータ量を削減することができる。すなわち、全てのグラフデータをメモリに格納することができるため、処理の高速化が可能となり、また、データ量の削減によって処理コストを抑制することができる。

#### 【実施例1】

40

#### 【0026】

図1は、本発明の実施例1のグラフ処理装置100の構成例を示すブロック図である。図2は、本発明の実施例1のグラフ処理装置100が適用されるシステム構成の一例を示すブロック図である。

#### 【0027】

図2に示すシステムは、グラフ処理装置100、基地局200、ユーザ端末210、及びセンサ群220から構成される。

#### 【0028】

グラフ処理装置100、基地局200及びセンサ群220に含まれる複数のセンサ221は、ネットワーク240を介して互いに接続される。ネットワーク240は、例えば、

50

WAN、LAN等が考えられるが、本発明はネットワーク240の種別に限定されない。

【0029】

ユーザ端末210は、基地局200を介して無線通信を介して、グラフ処理装置100等と接続される。なお、ユーザ端末210と基地局200との間には有線通信を介して接続されてよいし、ユーザ端末210が直接ネットワーク240と接続されてもよい。

【0030】

グラフ処理装置100は、センサ群220に含まれる各センサ221から業務データ130を取得し、取得された業務データ130をストレージ装置104に格納する。また、グラフ処理装置100は、ユーザ端末210の指示に従って、グラフ処理を実行する。

【0031】

ユーザ端末210は、例えばパーソナルコンピュータ又はタブレット端末等の装置である。ユーザ端末210は、プロセッサ(図示省略)、メモリ(図示省略)、ネットワークインタフェース(図示省略)、及び入出力装置(図示省略)を備える。入出力装置には、ディスプレイ、キーボード、マウス、及びタッチパネル等が含まれる。

【0032】

ユーザ端末210は、グラフ処理装置100を操作するためのユーザインタフェース211を提供する。ユーザインタフェース211は、グラフ処理装置100に目標処理時間を入力し、また、グラフ処理装置100から出力されたグラフデータ及びグラフ処理の結果等を受け付ける。

【0033】

グラフ処理装置100は、ハードウェア構成としてプロセッサ101、メモリ102、ネットワークインタフェース103、及びストレージ装置104を備える。

【0034】

プロセッサ101は、メモリ102に格納されるプログラムを実行する。プロセッサ101がプログラムを実行することによって、グラフ処理装置100が有する各種機能部を実現できる。以下の説明では、機能部を主体に処理の説明をするときには、当該機能部を実現するプログラムがプロセッサ101によって実行されていることを示す。

【0035】

メモリ102は、プロセッサ101によって実行されるプログラム及び当該プログラムの実行時に用いられる情報を格納する。メモリ102は、DRAM等が考えられる。メモリ102に格納されるプログラム及び情報については後述する。ネットワークインタフェース103は、WAN、LAN等のネットワークを介して外部の装置と接続するためのインタフェースである。

【0036】

ストレージ装置104は、各種情報を格納する。ストレージ装置104は、HDD又はSSD等が考えられる。本実施例では、ストレージ装置104に業務データ130が格納される。なお、業務データ130における各種データの相関関係を示す相関行列データが格納されてもよい。

【0037】

ここで、図3及び図4を用いて業務データ130及び相関行列データ400の一例について説明する。

【0038】

図3は、本発明の実施例1における業務データ130の一例を示す説明図である。図4は、本発明の実施例1における相関行列データ400の一例を示す説明図である。

【0039】

図3には、店舗における業務データ130を示す。業務データ130には、顧客毎の購入金額、購入点数、滞在時間、立ち止り時間などの情報が格納される。「購入金額」、「購入点数」、「滞在時間」、及び「立ち止り時間」を指標と呼ぶ。

【0040】

相関行列データ400は、指標間の相関関係を要素とする行列データである。例えば、

10

20

30

40

50



本実施例の行列データには、指標 1「購入金額」と指標 2「購入点数」との間の相関関係を示す情報が要素として含まれる。ここで、指標 1 と指標 2 との間の相関関係は、相関値として与えられる。例えば、下式 (1) を用いて相関値が算出される。

【 0 0 4 1 】

【 数 1 】

$$\frac{S_{12}}{S_1 \times S_2} \dots (1)$$

【 0 0 4 2 】

ここで、 $S_1$  は指標 1 の標準偏差、 $S_2$  は指標 2 の標準偏差、 $S_{12}$  は指標 1 と指標 2 との間の共分散を表す。

【 0 0 4 3 】

相関値は、「-1」以上「1」以下であり、相関値が「1」に近いほど「正の相関」が強いことを表し、相関値が「-1」に近いほど「負の相関」が強いことを表す。また、「0」に近いほど指標間に相関がないことを表す。

【 0 0 4 4 】

すなわち、相関行列データ 400 は、全ての指標の組合せに対する相関値を要素とする行列形式のデータ構造であり、指標間の関係性を示すデータである。以下の説明では、業務データ 130 から算出された相関行列データ 400 が、予め、ストレージ装置 104 に格納されるものとする。

【 0 0 4 5 】

図 1 の説明に戻る。次に、メモリ 102 に格納されるプログラム及び情報について説明する。

【 0 0 4 6 】

メモリ 102 には、グラフ処理部 110 を実現するプログラムを格納する。グラフ処理部 110 は、相関行列データ 400 をグラフデータに変換し、すなわち、相関行列データ 400 からグラフデータを生成する。また、グラフ処理部 110 は、グラフデータを用いて任意のグラフ処理を実行する。グラフ処理部 110 は、複数のプログラムモジュールから構成される。具体的には、グラフ処理部 110 は、エッジ情報量算出部 111、制御因子算出部 112、グラフデータ生成部 113、グラフ処理部 114、グラフデータ格納部 115 を含む。

【 0 0 4 7 】

エッジ情報量算出部 111 は、ストレージ装置 104 から相関行列データ 400 の要素を読み出し、相関値とエッジ数との間の関係を示すエッジ情報量を算出する。また、エッジ情報量算出部 111 は、算出されたエッジ情報量を制御因子算出部 112 に出力する。ここで、エッジ情報量は、相関行列データ 400 をグラフデータに変換する場合に含めることが可能なエッジの数を推定するための情報である。エッジ情報量算出部 111 が実行する処理の詳細は、図 6 を用いて後述する。

【 0 0 4 8 】

制御因子算出部 112 は、相関行列データ 400 をグラフデータに変換する場合に、データの圧縮に用いられる制御因子を算出する。本実施例では、制御因子算出部 112 は、エッジ情報量及び目標処理時間に基づいてグラフデータに含めるエッジ数を調整するための閾値を制御因子として算出する。また、制御因子算出部 112 は、算出された制御因子をグラフデータ生成部 113 に出力する。制御因子算出部 112 が実行する処理の詳細は、図 8 を用いて後述する。

【 0 0 4 9 】

グラフデータ生成部 113 は、算出された制御因子を用いて相関行列データ 400 からグラフデータを生成する。グラフデータ生成部 113 は、グラフデータ格納部 115 に生成されたグラフデータを格納し、また、ユーザ端末 210 に生成されたグラフデータを送

10

20

30

40

50

信する。グラフデータ生成部 113 が実行する処理の詳細は、図 11 を用いて後述する。

【0050】

グラフ処理部 114 は、グラフデータを用いて任意のグラフ処理を実行する。グラフ処理としては、例えば、行列演算の固有値計算に利用可能な Page Rank 処理、中心性計算処理等が考えられる。本発明は、グラフ処理の処理内容に限定されず、汎用的に用いられる様々なグラフアルゴリズムを適用することができる。グラフ処理部 114 は、ユーザ端末 210 にグラフ処理の結果を送信する。

【0051】

次に、本実施例のグラフ処理装置 100 が実行する処理について説明する。図 5 は、本発明の実施例 1 のグラフ処理装置 100 が実行する処理の概要を説明するフローチャートである。

10

【0052】

グラフ処理装置 100 は、ユーザ端末 210 から処理の開始指示を受信した場合、又は周期的に、以下で説明する処理を実行する。

【0053】

グラフ処理装置 100 は、ストレージ装置 104 に格納される業務データ 130 から相関行列データ 400 を生成する（ステップ S501）。具体的には、グラフ処理部 110 が相関行列データ 400 を生成する。なお、ストレージ装置 104 に相関行列データ 400 が格納されている場合には、ステップ S501 の処理は省略することができる。

【0054】

20

グラフ処理装置 100 は、エッジ情報量算出処理を実行する（ステップ S502）。具体的には、エッジ情報量算出部 111 が、相関行列データ 400 を解析し、解析結果に基づいてエッジ情報量を算出する。エッジ情報量算出部 111 が実行するエッジ情報量算出処理の詳細は、図 6 を用いて後述する。

【0055】

グラフ処理装置 100 は、ユーザ端末 210 から目標処理時間を取得する（ステップ S503）。具体的には、グラフ処理部 110 が、ユーザ端末 210 に対して目標処理時間の入力を要求する。このとき、ユーザインタフェース 211 は、当該要求を受け付けると、ディスプレイ等に目標処理時間を入力するための操作画面を表示し、当該操作画面を用いて入力された目標処理時間をグラフ処理装置 100 に送信する。グラフ処理装置 100 は、ユーザ端末 210 から受信した目標処理時間を制御因子算出部 112 に入力する。

30

【0056】

グラフ処理装置 100 は、エッジ情報量及び目標処理時間を用いて制御因子算出処理を実行する（ステップ S504）。具体的には、制御因子算出部 112 が、エッジ情報量及び目標処理時間を用いて、圧縮されたグラフデータの生成に用いる制御因子を算出する。制御因子算出部 112 が実行する制御因子算出処理の詳細は、図 8 を用いて後述する。

【0057】

グラフ処理装置 100 は、制御因子を用いてグラフデータ生成処理を実行する（ステップ S505）。具体的には、グラフデータ生成部 113 が、算出された制御因子を用いて、相関行列データ 400 からグラフデータを生成する。グラフデータ生成部 113 が実行するグラフデータ生成処理の詳細は、図 11 を用いて後述する。

40

【0058】

グラフ処理装置 100 は、生成されたグラフデータを用いてグラフ処理を実行する（ステップ S506）。具体的には、グラフ処理部 114 が、生成されたグラフデータを用いて所定のグラフ処理を実行し、グラフ処理の結果をユーザ端末 210 に送信する。

【0059】

図 6 は、本発明の実施例 1 のエッジ情報量算出処理の一例を説明するフローチャートである。図 7 A は、本発明の実施例 1 の相関値の頻度分布表 700 の一例を示す説明図である。図 7 B は、本発明の実施例 1 のエッジ情報量の一例を示す説明図である。

【0060】

50

エッジ情報量算出部 111 は、相関行列データ 400 における相関値の頻度分布表（ヒストグラム）700 を生成する（ステップ S601）。

【0061】

ここで、相関値の頻度分布表 700 は、相関値を所定の値の範囲毎にカウントした度数分布を表す柱状グラフであり、図 7A に示すようなグラフとなる。図 7A では、値の範囲は「0.01」である。なお、相関値の頻度分布表 700 における値の範囲は予め設定されているものとする。ただし、外部からの入力に基づいて、値の範囲を変更することができる。

【0062】

エッジ情報量算出部 111 は、相関行列データ 400 の要素のループ処理を開始する（ステップ S602）。まず、エッジ情報量算出部 111 は、相関行列データ 400 から要素を一つ選択し、選択された要素の値（相関値）を読み出す。

10

【0063】

エッジ情報量算出部 111 は、読み出された要素の値の絶対値、すなわち、相関値の絶対値を算出する（ステップ S603）。エッジ情報量算出部 111 は、算出された相関値の絶対値に基づいて、相関値の頻度分布表 700 を更新する（ステップ S604）。具体的には、エッジ情報量算出部 111 は、相関値の絶対値が含まれる値の範囲の度数を 1 加算する。なお、エッジ情報量算出部 111 は、相関値の頻度分布表 700 の更新後、読み出された要素の値を削除する。

【0064】

20

エッジ情報量算出部 111 は、相関行列データ 400 の全ての要素について処理が完了したか否かを判定する（ステップ S605）。相関行列データ 400 の全ての要素について処理が完了していないと判定された場合、エッジ情報量算出部 111 は、ステップ S602 に戻り、同様の処理を実行する。一方、相関行列データ 400 の全ての要素について処理が完了したと判定された場合、エッジ情報量算出部 111 は、ステップ S606 に進む。

【0065】

相関行列データ 400 の要素のループ処理が完了すると、相関値の頻度分布表 700 は図 7A に示すような状態になる。

【0066】

30

エッジ情報量算出部 111 は、相関値の頻度分布表 700 に基づいてエッジ情報量を算出し（ステップ S606）、制御因子算出部 112 に算出されたエッジ情報量を出力する（ステップ S607）。その後、エッジ情報量算出部 111 は処理を終了する。具体的には、以下のような処理が実行される。

【0067】

エッジ情報量算出部 111 は、相関値の絶対値「k」までの度数の合計値、すなわち、度数の累積頻度を算出する。横軸を相関値の絶対値、縦軸を度数の累積頻度として、算出された度数の累積頻度をプロットする。エッジ情報量算出部 111 は、プロット結果から相関値の絶対値と累積頻度との間の関係を示す関数  $E(k)$  をエッジ情報量として算出する。本実施例では、エッジ情報量  $E(k)$  は、図 7B に示すようなグラフ 701 として与えられる。

40

【0068】

累積頻度は、相関値の頻度分布表 700 における相関値の絶対値が「k」までの度数の合計値を表す。例えば、 $E(0.3)$  は相関値の絶対値が「0」から「0.3」までの度数の合計値である。したがって、 $E(1)$  は相関行列データ 400 の全要素の数と一致する。

【0069】

図 8 は、本発明の実施例 1 の制御因子算出処理の一例を説明するフローチャートである。図 9 は、本発明の実施例 1 の推定処理時間関数  $f(E)$  の一例を示す説明図である。図 10 は、本発明の実施例 1 の制御因子の決定時に用いられる推定用エッジ情報量の一例を

50

示す説明図である。

【0070】

制御因子算出部112は、エッジ情報量が入力されると処理を開始する。制御因子算出部112は、エッジ情報量 $E(k)$ を変数とする推定処理時間関数 $f(E)$ を求める(ステップS801)。

【0071】

制御因子算出部112は、グラフ解析処理のアルゴリズムに基づいて推定処理時間関数 $f(E)$ を算出することができる。例えば、グラフ分析処理において、主成分分析に用いる固有値問題を解く場合、アルゴリズムの収束計算の繰り返し回数を $a$ 、単位エッジ当たりの処理時間を $b$ 、変数 $E$ として場合、下式(2)で与えられる。

【0072】

【数2】

$$f(E) = a \times b \times E \quad \dots (2)$$

【0073】

図9には、式(2)によって求められた推定処理時間関数 $f(E)$ を示す。なお、エッジ情報量 $E(k)$ は推定処理時間関数 $f(E)$ の定義域として与えられる。

【0074】

次に、制御因子算出部112は、ユーザ端末210から目標処理時間を取得する(ステップS802)。例えば、制御因子算出部112は、ユーザ端末210に対して、目標処理時間の入力を要求する。ユーザ端末210は、ユーザインタフェース211を介して当該要求を受け付けると、ディスプレイに目標処理時間を入力するための操作画面等を表示する。以下の説明では取得された目標処理時間が $T$ であるものとする。

【0075】

制御因子算出部112は、目標処理時間及び推定処理時間関数 $f(E)$ を用いて、目標処理時間内にグラフ処理が完了可能な最大エッジ数 $E_{MAX}$ を算出する(ステップS803)。

【0076】

本実施例では、制御因子算出部112は式(2)から最大エッジ数 $E$ を算出できる。具体的には下式(3)のように最大エッジ数 $E_{MAX}$ が算出される。図9の点線は、式(3)を用いて算出される最大エッジ数 $E_{MAX}$ を示す。

【0077】

【数3】

$$E_{MAX} = \frac{T}{a \times b} \quad \dots (3)$$

【0078】

制御因子算出部112は、エッジ情報量 $E(k)$ 及び最大エッジ数 $E_{MAX}$ を用いて相関値の閾値を算出する(ステップS804)。具体的には、以下のような処理が実行される。

【0079】

制御因子算出部112は、まず、エッジ情報量 $E(k)$ を用いて推定用エッジ情報量 $E'(k)$ を求める。本実施例では、下式(4)に示すように推定用エッジ情報量 $E'(k)$ が求められる。推定用エッジ情報量 $E'(k)$ は、図10に示すようなグラフ1000として与えられる。

【0080】

【数4】

$$E'(k) = E(1) - E(k) \quad \dots (4)$$

【0081】

10

20

30

40

50

制御因子算出部 112 は、推定用エッジ情報量  $E'(k)$  及び最大エッジ数  $E_{MAX}$  を用いて相関値の閾値を算出する。具体的には、制御因子算出部 112 は、式 (4) の左辺を  $E_{MAX}$  とし、下式 (5) のように変更することによって相関値の絶対値  $k$  を算出する。算出された相関値の絶対値  $k$  が相関値の閾値となる。図 10 の点線は、式 (5) を用いて算出された相関値の閾値を示す。相関値の閾値は、後述するように、グラフデータ生成処理において相関値の切り捨ての閾値 (制御因子) として用いられる。

【0082】

【数5】

$$E(k) = E(1) - E_{MAX} \quad \dots (5)$$

10

【0083】

制御因子算出部 112 は、グラフデータ生成部 113 に、算出された相関値の閾値を制御因子として出力し (ステップ S805)、処理を終了する。

【0084】

図 11 は、本発明の実施例 1 のグラフデータ生成処理の一例を説明するフローチャートである。図 12A は、本発明の実施例 1 のグラフデータ生成処理に用いられる頂点リスト 1200 の一例を示す説明図である。図 12B は、本発明の実施例 1 のグラフデータ生成処理に用いられるエッジリスト 1210 の一例を示す説明図である。図 13 は、本発明の実施例 1 のグラフデータ生成処理における制御因子を用いた相関値の切り捨ての概念を示す説明図である。図 14A 及び図 14B は、本発明の実施例 1 のグラフデータ生成処理の実行後の頂点リスト 1200 及びエッジリスト 1210 を示す説明図である。図 15 は、本発明の実施例 1 のグラフデータに基づいて表示されるグラフの一例を示す説明図である。

20

【0085】

まず、頂点リスト 1200 及びエッジリスト 1210 について説明する。

【0086】

頂点リスト 1200 は、グラフデータにおける頂点 (指標)、及び頂点を接続するエッジの情報を管理するための情報である。図 12A に示す頂点リスト 1200 は、頂点 ID 1201、指標 ID 1202、及び接続エッジ情報 1203 を含む。

【0087】

頂点 ID 1201 は、頂点を一意に識別するための識別情報を格納する。一つの頂点に対して一つの頂点 ID が付与される。指標 ID 1202 は、頂点に対応する指標の識別情報である。グラフデータでは、一つの指標が一つの頂点として管理される。接続エッジ情報 1203 は、頂点 ID 1201 に対応する頂点に接続されるエッジの情報である。

30

【0088】

エッジリスト 1210 は、グラフデータにおけるエッジ (辺) を管理するための情報である。図 12B に示すエッジリスト 1210 はエッジ ID 1211、接続頂点 A 1212、接続頂点 B 1213、及び重み 1214 を含む。

【0089】

エッジ ID 1211 は、エッジを一意に識別するための識別情報を格納する。一つのエッジに対して一つのエッジ ID が付与される。接続頂点 A 1212 及び接続頂点 B 1213 は、エッジによって接続される二つの頂点の識別情報を格納する。重み 1214 は、エッジの重み、すなわち、相関値を格納する。

40

【0090】

グラフデータ生成部 113 は、制御因子が入力されると処理を開始する。グラフデータ生成部 113 は、まず、頂点リスト 1200 及びエッジリスト 1210 を初期化する (ステップ S1101)。

【0091】

具体的には、グラフデータ生成部 113 は、相関行列データ 400 の全ての指標の数だけ頂点リスト 1200 にエントリを生成し、生成されたエントリの指標 ID 1202 に指

50

標の識別情報を設定する。グラフデータ生成部 113 は、各指標に頂点 ID を付与し、各エントリの頂点 ID 1201 に付与された頂点 ID を設定する。この時点では、接続エッジ情報 1203 は空の状態である。また、グラフデータ生成部 113 は、空のエッジリスト 1210 を生成する。

【0092】

グラフデータ生成部 113 は、相関行列データ 400 の要素のループ処理を開始する（ステップ S1102）。まず、グラフデータ生成部 113 は、相関行列データ 400 から要素を一つ読み出す。なお、グラフデータ生成部 113 が要素を一つずつ読み出すと、頻繁に I/O が発生するため、例えば、相関行列データ 400 の行単位に要素を読み出し、読み出された要素をメモリ 102 に一時的に保持してもよい。

10

【0093】

グラフデータ生成部 113 は、読み出された要素の相関値の絶対値が相関値の閾値（制御因子）より小さいか否かを判定する（ステップ S1103）。読み出された要素の相関値の絶対値が相関値の閾値（制御因子）より小さいと判定された場合、グラフデータ生成部 113 は、ステップ S1105 に進む。

【0094】

読み出された要素の相関値の絶対値が相関値の閾値（制御因子）以上であると判定された場合、グラフデータ生成部 113 は、頂点リスト 1200 及びエッジリスト 1210 を更新する（ステップ S1104）。具体的には、以下のような処理が実行される。

【0095】

20

グラフデータ生成部 113 は、エッジリスト 1210 にエントリを追加し、追加されたエントリのエッジ ID 1211 にエッジの識別情報を設定する。また、グラフデータ生成部 113 は、追加されたエントリの接続頂点 A 1212 及び接続頂点 B 1213 に、読み出された要素に対応する二つの指標を設定する。さらに、グラフデータ生成部 113 は、追加されたエントリの重み 1214 に読み出された要素の相関値を設定する。

【0096】

グラフデータ生成部 113 は、頂点リスト 1200 を参照し、指標 ID 1202 が接続頂点 A 1212 に設定された指標の識別情報と一致するエントリを検索する。グラフデータ生成部 113 は、検索されたエントリの接続エッジ情報 1203 に、エッジ ID 1211 に設定されたエッジの識別情報を設定する。グラフデータ生成部 113 は、同様に、指標 ID 1202 が接続頂点 B 1213 に設定された指標の識別情報と一致するエントリを検索し、当該エントリの接続エッジ情報 1203 にエッジの識別情報を設定する。

30

【0097】

なお、接続エッジ情報 1203 に、追加予定のエッジの識別情報と同一のエッジの識別情報が格納されている場合、グラフデータ生成部 113 は、追加予定のエッジの識別情報を設定しない。これは追加する必要がないためである。

【0098】

以上がステップ S1104 の処理の説明である。

【0099】

グラフデータ生成部 113 は、相関行列データ 400 の全ての要素について処理が完了したか否かを判定する（ステップ S1105）。相関行列データ 400 の全ての要素について処理が完了していないと判定された場合、グラフデータ生成部 113 は、ステップ S1102 に戻り、同様の処理を実行する。一方、相関行列データ 400 の全ての要素について処理が完了したと判定された場合、グラフデータ生成部 113 は、ステップ S1106 に進む。

40

【0100】

相関行列データ 400 の要素のループ処理は、図 13 に示すように、相関値の絶対値が相関値の閾値（制御因子）より小さい要素の値を「0」に設定して、その後グラフデータを生成する処理に対応する。

【0101】

50

グラフデータ生成部 113 は、頂点リスト 1200 を参照し、いずれのエッジにも接続されていない頂点のエントリを当該頂点リスト 1200 から削除する（ステップ S1106）。具体的には、グラフデータ生成部 113 は、接続エッジ情報 1203 にエッジの識別情報が一つも格納されていないエントリを検索し、当該エントリを頂点リスト 1200 から削除する。

【0102】

以上の処理が終了すると、頂点リスト 1200 及びエッジリスト 1210 は、図 14A 及び 図 14B に示すような状態になる。

【0103】

グラフデータ生成部 113 は、頂点リスト 1200 及びエッジリスト 1210 をグラフデータとして出力し（ステップ S1107）、処理を終了する。本実施例では、グラフデータ生成部 113 は、頂点リスト 1200 及びエッジリスト 1210 をグラフデータ格納部 115 に出力し、また、ユーザ端末 210 に送信する。ユーザ端末 210 は、受信したグラフデータに基づいて 図 15 に示すようなグラフを表示することができる。

【0104】

本実施例では、グラフデータは、頂点リスト 1200 及びエッジリスト 1210 から構成されるものとするが、本発明はリスト表現に限定されず、そのほかのグラフ表現方法を用いてもよい。

【0105】

ここで、図 4、図 14A、図 14B、及び 図 15 を用いて相関行列データ 400 とグラフデータとのデータ量について説明する。

【0106】

図 4 に示すように、5 行 5 列の相関行列データ 400 では、25 個の指標の組合せのそれぞれについて相関値を保持する必要がある。一方、グラフデータでは、5 個の頂点の情報と、エッジの重みを含む 10 個のエッジの情報とを保持すればよい。したがって、グラフ処理装置 100 は、相関行列データ 400 をグラフデータに変換することによって、データ量を圧縮することができる。

【0107】

実施例 1 によれば、グラフ処理装置 100 は、単に相関行列データ 400 をグラフデータに変換するだけでなく、目標処理時間内に処理が完了できるように制御因子を用いてグラフデータに含まれるエッジの数を調整し、その後、グラフデータを生成する。これによって、生成されたグラフデータは更に圧縮されたデータとなるため、メモリ 102 にデータを配置することができ、当該メモリ 102 上のグラフデータを用いて高速なグラフ解析処理が可能となる。すなわち、相関行列データをグラフデータとして圧縮し、大量の指標の相関分析又は主成分分析等のビックデータ解析において、データ量を削減し、かつ、高速な処理を実現できる。

【0108】

（変形例）

実施例 1 では、相関値の絶対値が相関値の閾値より小さい要素の値を「0」とすることによってエッジとして保持するデータ量を削減したが、本発明はこれに限定されない。例えば、グラフデータ生成部 113 は、相関値の絶対値が相関値の閾値より大きい要素のみを抽出し、抽出された要素からグラフデータを生成してもよい。

【実施例 2】

【0109】

次に実施例 2 について説明する。実施例 2 では、目標処理時間だけではなく、ユーザによって指定されたメモリ制限をも考慮して、さらに、圧縮されたグラフデータを生成する。具体的には、制御因子算出部 112 が、グラフデータに含めるエッジ数を調整するため閾値、及びエッジの重みの表現ビット数を制御因子として算出する。これによって、グラフ処理装置 100 は、エッジ数の削減し、さらに、エッジの重みの表現ビット数を丸めることによって、さらに、データ量を圧縮する。以下実施例 1 との差異を中心に実施例 2

10

20

30

40

50

について説明する。なお、実施例 1 と同一の構成には同一の符号を付し、詳細な説明は省略する。

【0110】

図 16 は、本発明の実施例 2 のグラフ処理装置 100 の構成例を示すブロック図である。なお、グラフ処理装置 100 が適用されるシステム構成例は実施例 1 と同一であるため説明を省略する。

【0111】

図 16 に示すように、ユーザ端末 210 は、目標処理時間に加え、メモリ制限量を入力する点が発明例 1 と異なる。制御因子算出部 112 は、目標処理時間及びメモリ制限量に基づいて、相関値の閾値及びエッジの重みに対する丸めビット数を算出する。その他の構成は実施例 1 と同一である。

10

【0112】

相関行列データ 400 のデータ形式は、実施例と同一であるため説明を省略する。グラフ処理装置 100 が実行する処理の概要も実施例 1 と同一であるため説明を省略する。また、エッジ情報量算出処理も実施例 1 と同一であるため説明を省略する。実施例 2 では、制御因子算出処理及びグラフデータ生成処理の一部の内容が異なる。

【0113】

図 17 は、本発明の実施例 2 の制御因子算出処理の一例を説明するフローチャートである。図 18A 及び図 18B は、本発明の実施例 2 の推定メモリ使用量関数  $g(E, B)$  の一例を示す説明図である。図 19 は、本発明の実施例 2 の相関値の表現ビット数の丸めの一例を示す説明図である。

20

【0114】

実施例 2 の制御因子算出処理では、制御因子算出部 112 は、推定処理時間関数  $f(E)$  を求めた後、相関値の表現ビット数毎に、エッジ情報量に対する推定メモリ使用量関数  $g(E, B)$  を求める (ステップ S1701)。ここで、 $E$  はエッジ数、 $B$  は表現ビット数を表す。

【0115】

推定メモリ使用量関数  $g(E, B)$  はエッジの重みを何ビットで表現するかによって複数存在する。例えば、1 ビットで重みを表現した場合の一つのエッジ当たりのメモリ使用量を  $x$ 、エッジ数を  $E$ 、エッジのビット数  $y$  として場合、推定メモリ使用量関数  $g(E, B)$  は、下式 (6) ように求められる。

30

【0116】

【数 6】

$$g(E, y) = x \times y \times E \quad \dots (6)$$

【0117】

図 18A 及び図 18B には、式 (6) によって求められた推定メモリ使用量関数  $g(E, B)$  を示す。なお、エッジ情報量  $E(k)$  は推定メモリ使用量関数  $g(E, B)$  の定義域として与えられる。

【0118】

ステップ S1701 の後、制御因子算出部 112 は、ユーザ端末 210 から目標処理時間及びメモリ制限量を取得する (ステップ S1702)。メモリ制限量の取得方法は、目標処理時間と同様の方法を用いればよい。以下の説明では取得された目標処理時間が  $T$ 、メモリ制限量が  $G$  であるものとする。

40

【0119】

制御因子算出部 112 は、最大エッジ数  $E_{MAX}$  を算出した後 (ステップ S803)、最大エッジ数、メモリ制限量、及び推定メモリ使用量関数  $g(E, B)$  に基づいて、エッジの重みの表現ビット数を決定する (ステップ S1703)。具体的には、以下のような処理が実行される。

【0120】

50



制御因子算出部 112 は、各推定メモリ使用量関数  $g(E, B)$  に最大エッジ数  $E_{MAX}$  を代入し、推定メモリ使用量を算出する。制御因子算出部 112 は、算出された推定メモリ使用量が下式(7)を満たすものを抽出する。

【0121】

【数7】

$$g(E_{MAX}, B) \leq G \quad \dots (7)$$

【0122】

制御因子算出部 112 は、式(7)を満たす推定メモリ使用量の中から最も大きいビット数を特定し、特定されたビット数をエッジの重みの表現ビット数に決定する。

10

【0123】

例えば、図18Aに示す例ではエッジの重みの表現ビット数は3ビットと決定され、図18Bに示す例ではエッジの重みの表現ビット数は2ビットと決定される。

【0124】

制御因子算出部 112 は、相関値の閾値を算出した後(ステップS804)、グラフデータ生成部 113 に、当該相関値の閾値及び表現ビット数を制御因子として出力し(ステップS1704)、処理を終了する。

【0125】

実施例2のグラフ生成処理の流れは、実施例1のグラフ生成処理(図11参照)と同一である。ただし、ステップS1104の処理が一部異なる。

20

【0126】

具体的には、グラフデータ生成部 113 は、エッジリスト 1210 に追加されたエントリの重み 1214 に相関値を設定する場合、制御因子として入力された表現ビット数に基づいて相関値を丸めて、丸められた相関値を重み 1214 に設定する。

【0127】

例えば、丸める前の相関値の表現ビット数が4ビットであり、これを3ビットに丸める場合、最上位ビットは符号ビットとする。例えば、「0」の場合「正」の相関値に対応し、「1」の場合「負」の相関値に対応するようにすればよい。また、相関値の絶対値の大きさに応じて図19に示すような符号化を与えればよい。なお、符号は図19に示す以外の符号化であってもよい。

30

【0128】

その他の処理は実施例1と同一である。

【0129】

実施例2によれば、メモリ制限に従って、エッジの重みの表現ビット数を丸めることによって、グラフデータを更に圧縮することができる。すなわち、システムにおいて使用可能なメモリ容量の制約のもと、目標処理時間内に処理可能なデータ量のグラフデータを生成することができる。これによって、相関行列データ 400 から生成されたグラフデータを全てメモリ 102 上に配置し、メモリ 102 上に配置されたデータを用いて高速なグラフ処理が可能となる。

【0130】

40

なお、本発明は上記した実施例に限定されるものではなく、様々な変形例が含まれる。また、例えば、上記した実施例は本発明を分かりやすく説明するために構成を詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。また、各実施例の構成の一部について、他の構成に追加、削除、置換することが可能である。

【0131】

また、上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、本発明は、実施例の機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体をコンピュータに提供し、そのコンピュータが備えるブ

50

ロセッサが記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施例の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク、CD-ROM、DVD-ROM、ハードディスク、SSD (Solid State Drive)、光ディスク、光磁気ディスク、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどが用いられる。

【0132】

また、本実施例に記載の機能を実現するプログラムコードは、例えば、アセンブラ、C/C++、perl、Shell、PHP、Java (登録商標) 等の広範囲のプログラム又はスクリプト言語で実装できる。

10

【0133】

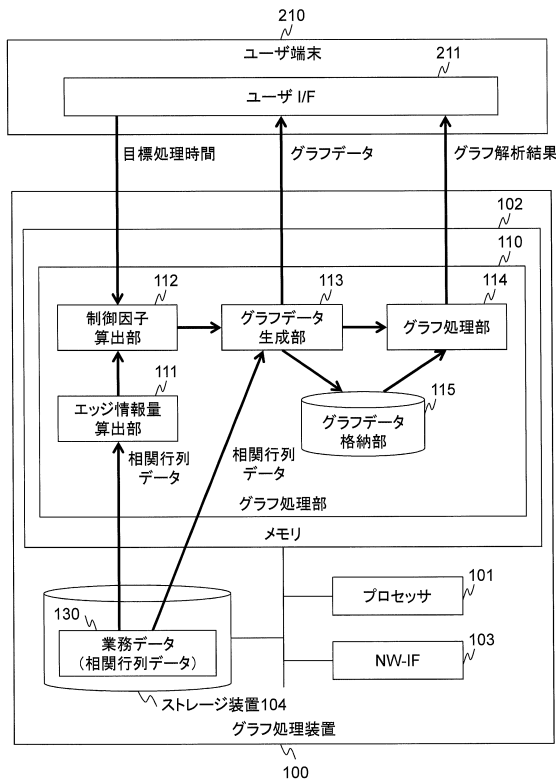
さらに、実施例の機能を実現するソフトウェアのプログラムコードを、ネットワークを介して配信することによって、それをコンピュータのハードディスクやメモリ等の記憶手段又はCD-RW、CD-R等の記憶媒体に格納し、コンピュータが備えるプロセッサが当該記憶手段や当該記憶媒体に格納されたプログラムコードを読み出して実行するようにしてもよい。

【0134】

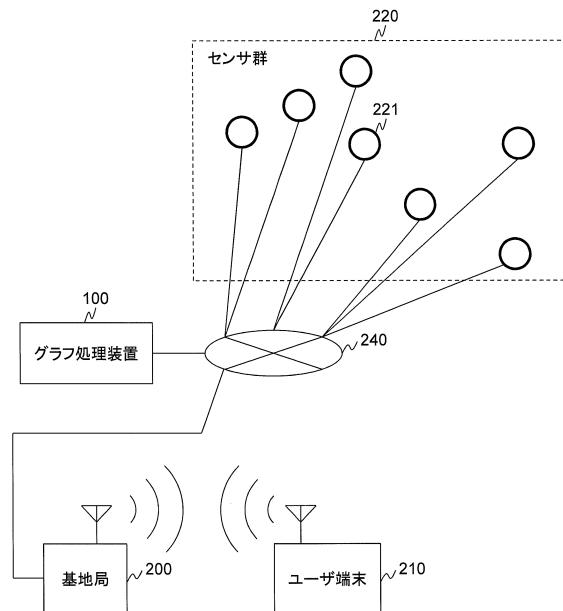
上述の実施例において、制御線や情報線は、説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。全ての構成が相互に接続されていてもよい。

20

【図1】



【図2】



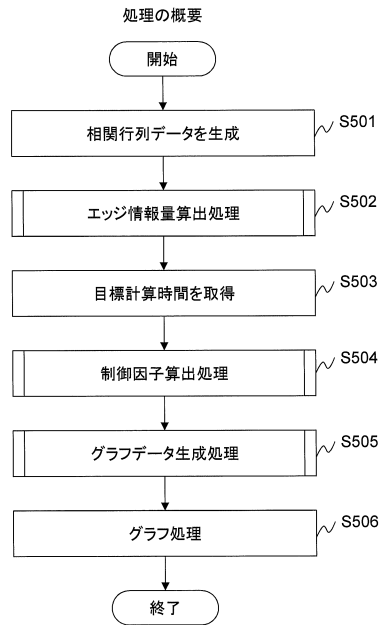
【図3】

	指標1	指標2	指標3	指標4	...
	購入金額 (円)	購入点数 (個)	滞在時間 (分)	立ち止まり時間 (分)	...
客1	1500	3	45	25	...
客2	1000	2	32	10	...
客3	3500	4	64	38	...
⋮	⋮	⋮	⋮	⋮	⋮

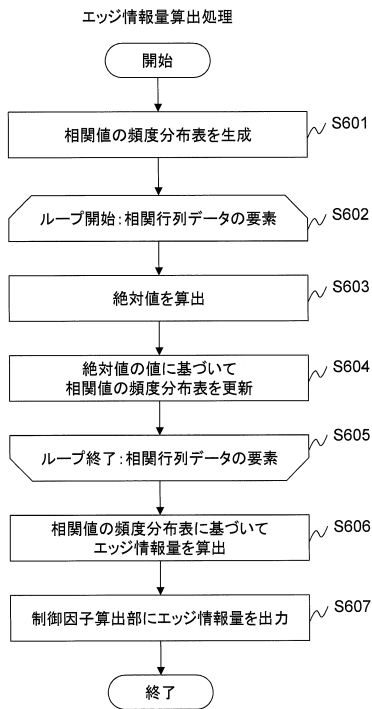
【図4】

	指標1	指標2	指標3	指標4	指標5
指標1	1.0	0.11	0.43	-0.12	0.13
指標2	0.11	1.0	0.56	0.45	0.08
指標3	0.43	0.56	1.0	-0.13	-0.25
指標4	-0.12	0.45	-0.13	1.0	0.48
指標5	0.13	0.08	-0.25	0.48	1.0

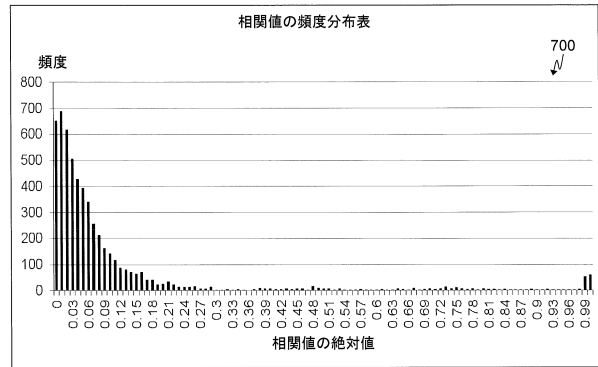
【図5】



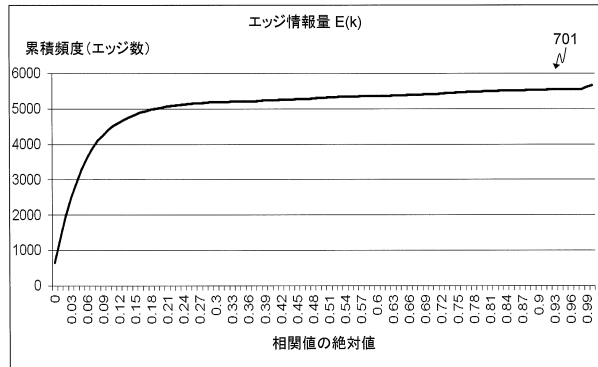
【図6】



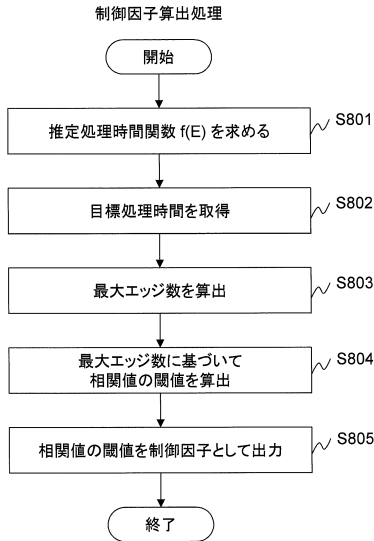
【図7A】



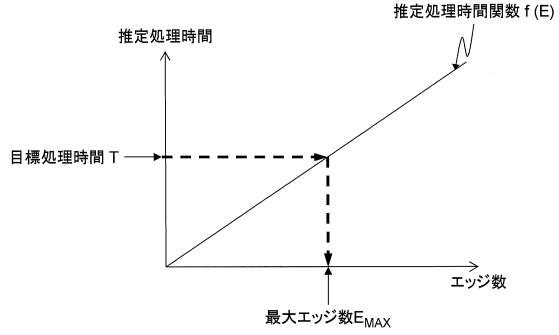
【図7B】



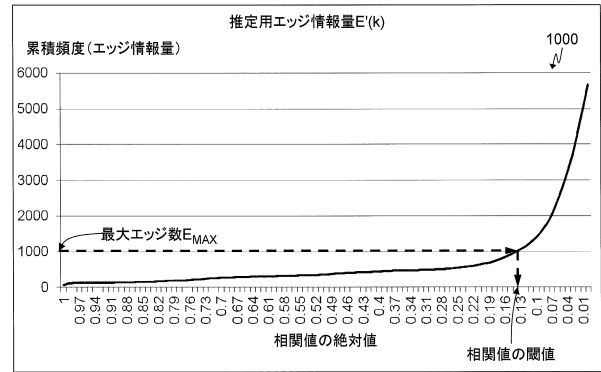
【図 8】



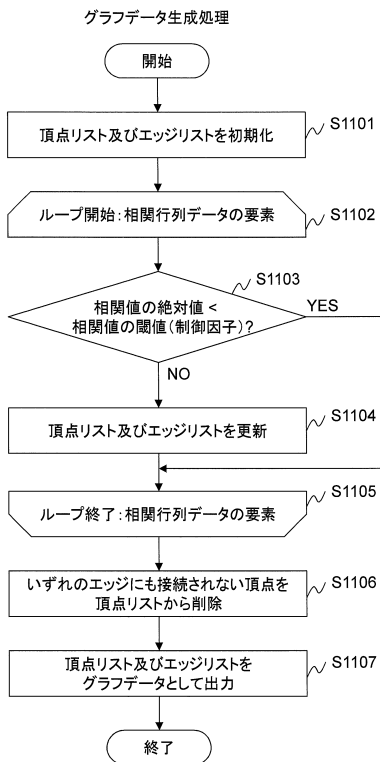
【図 9】



【図 10】



【図 11】



【図 12 A】

頂点リスト

1201	1202	1203	1200
頂点ID	指標 ID	接続エッジ情報	
頂点1	指標1		
頂点2	指標2		
頂点3	指標3		
頂点4	指標4		
頂点5	指標5		

【図 12 B】

エッジリスト

1211	1212	1213	1214	1210
エッジID	接続頂点A	接続頂点B	重み	

【図13】

	指標1	指標2	指標3	指標4	指標5
指標1	1.0	0.11	0.43	-0.12	0.13
指標2	0.11	1.0	0.56	0.45	0.08
指標3	0.43	0.56	1.0	-0.13	-0.25
指標4	-0.12	0.45	-0.13	1.0	0.48
指標5	0.13	0.08	-0.25	0.48	1.0



	指標1	指標2	指標3	指標4	指標5
指標1	1.0	0	0.43	0	0
指標2	0	1.0	0.56	0.45	0
指標3	0.43	0.56	1.0	0	-0.25
指標4	0	0.45	0	1.0	0.48
指標5	0	0	-0.25	0.48	1.0

【図14A】

頂点リスト

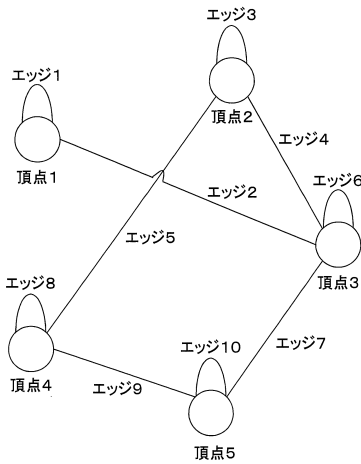
頂点ID	指標ID	接続エッジ情報
頂点1	指標1	エッジ1、エッジ2
頂点2	指標2	エッジ3、エッジ4、エッジ5
頂点3	指標3	エッジ2、エッジ4、エッジ6、エッジ7
頂点4	指標4	エッジ5、エッジ8、エッジ9
頂点5	指標5	エッジ7、エッジ9、エッジ10

【図14B】

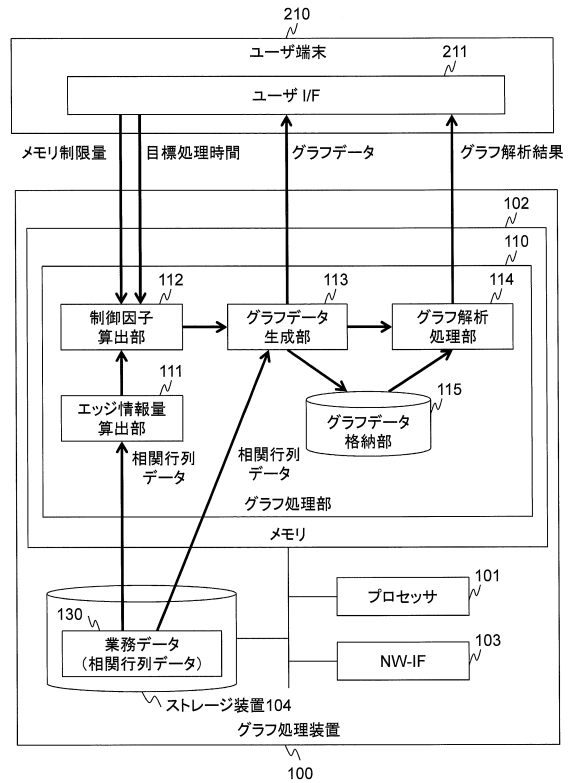
エッジリスト

エッジID	接続頂点A	接続頂点B	重み
エッジ1	頂点1	頂点1	1.0
エッジ2	頂点1	頂点3	0.43
エッジ3	頂点2	頂点2	1.0
エッジ4	頂点2	頂点3	0.56
エッジ5	頂点2	頂点4	0.45
エッジ6	頂点3	頂点3	1.0
エッジ7	頂点3	頂点5	-0.25
エッジ8	頂点4	頂点4	1.0
エッジ9	頂点4	頂点5	0.48
エッジ10	頂点5	頂点5	1.0

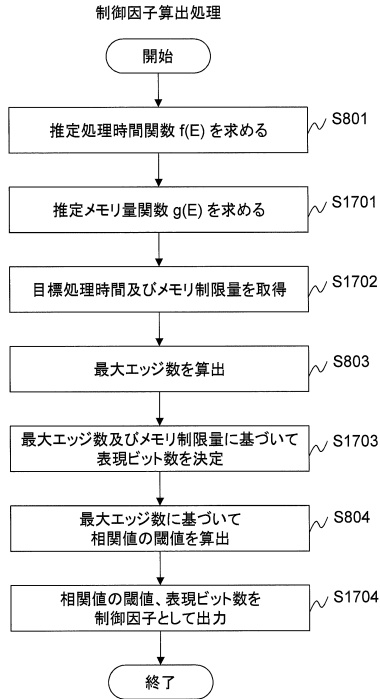
【図15】



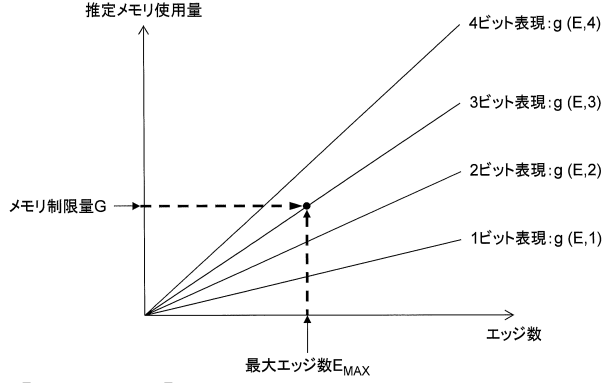
【図16】



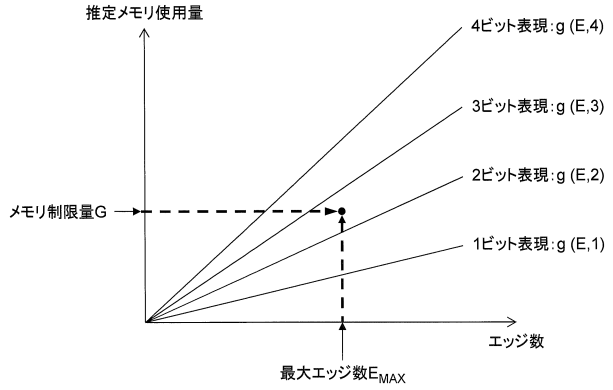
【図 17】



【図 18 A】



【図 18 B】



【図 19】

符号ビット	相関値の絶対値の範囲
00	0.0~0.25
01	0.25~0.5
10	0.5~0.75
11	0.75~1.0

---

フロントページの続き

審査官 田中 幸雄

(56)参考文献 特開2007-87125(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/18

G06F 17/16