

(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
2013 年 12 月 27 日(27.12.2013)



(10) 国際公開番号

WO 2013/190627 A1

- (51) 国際特許分類:
G06Q 50/04 (2012.01)
- (21) 国際出願番号: PCT/JP2012/065552
- (22) 国際出願日: 2012 年 6 月 18 日(18.06.2012)
- (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(米国を除く全ての指定国について): 株式会社日立製作所 (HITACHI, LTD.) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目 6 番 6 号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 森 靖英 (MORI, Yasuhide) [JP/JP]; 〒2440817 神奈川県横浜市戸塚区吉田町 2 9 2 番地 株式会社日立製作所 横浜研究所内 Kanagawa (JP). 野口 孝史 (NOGUCHI, Takashi) [JP/JP]; 〒2440817 神奈川県横浜市戸塚区吉田町 2 9 2 番地 株式会社日立製作所 横浜研究所内 Kanagawa (JP). 飯島 光一朗

(IJJIMA, Koichiro) [JP/JP]; 〒2440817 神奈川県横浜市戸塚区吉田町 2 9 2 番地 株式会社日立製作所 横浜研究所内 Kanagawa (JP).

(74) 代理人: 大賀 眞司, 外(OHGA, Shinji et al.); 〒1400002 東京都品川区東品川二丁目 3 番 1 2 号 シーフォートスクエア センタービルディング 1 6 階 サンネクスト国際特許事務所 Tokyo (JP).

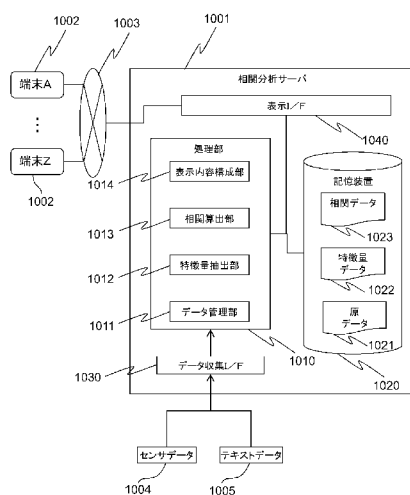
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: CORRELATION ANALYZING DEVICE AND CORRELATION ANALYZING METHOD

(54) 発明の名称: 相関分析装置及びその方法

図 1



1001 Correlation analyzing server
1002 Terminal A ... Terminal Z
1004 Sensor data
1005 Text data
1010 Processing unit
1011 Data management section
1012 Feature amount extracting section
1013 Correlation calculation section
1014 Display content configuration section
1020 Storage device
1021 Original data
1022 Feature amount data
1023 Correlation data
1030 Data collection interface
1040 Display interface

(57) Abstract: A correlation analyzing device includes a data conversion section for converting a plurality of data items generated in different events and having different types or different temporal resolutions to data having a common data format by using relationship information that manages the relationships among the plurality of data items, and a correlation calculation section for calculating the correlations among the data obtained by the conversion in the data conversion section.

(57) 要約: イベント毎に生起されるデータであって、種類または時間解像度の異なる複数のデータを、前記複数のデータ間の関連性を管理する関連性情報を用いて、各データ共通のデータ形式のデータに変換するデータ変換部と、前記データ変換部で変換された各データの相関を算出する相関算出部とを有する。



(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, 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, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称： 相関分析装置及びその方法

技術分野

[0001] 本発明は、イベント間の関連性を分析する装置、特に、複数種類のデータを用いてイベントの原因分析の支援を行う相関分析装置及びその方法に関する。

背景技術

[0002] 背景技術として、人手によるメソッドを定義したR C A (Root Cause Analysis) やF T A (Fault Tree Analysis) がある。前者は、原因を木構造でリストアップしながら何段階か辿って根本原因を追究する手法である。後者は、因果の関連性をANDやORの分岐を持った木構造で表現するものである。

[0003] 原因究明作業を支援するシステムの発明として、特許文献1がある。特許文献1には、単位時間当たりの変化が予め設定した閾値を超えた場合に、選別された他のデータを表示する機能を持つ技術が開示されている。

[0004] また、原因究明作業を支援するシステムの発明として、特許文献2がある。特許文献2には、逸脱事象の発生原因を分析する際に、ヒューマンエラー樹形図を参照して、分析者に、逸脱事象の発生原因の分析を支援する支援情報を呈示する機能を持つ技術が開示されている。

先行技術文献

特許文献

[0005] 特許文献1：特開2003-216238号公報

特許文献2：特開2004-287649号公報

発明の概要

発明が解決しようとする課題

[0006] 上記各特許文献に記載された発明では、扱える時系列の特徴量は均一であることを想定している。一方、原因を追求するための基礎データは、一般に複数種類のデータ（言語をデータ化したテキストデータ、センサデータや各

種計測データ等の数値データ）が混在している。そのため、複数種類の基礎データを用いて原因分析する際には、各基礎データ間の関係を比較する必要がある。しかし、この点については、従来技術では何ら配慮されていない。また、複数種類の基礎データを用いて原因分析する際に、基礎データを大規模データに適用するための計算量やデータ量を削減することについても、従来技術では何ら配慮されていない。

[0007] 本発明の目的は、複数種類のデータが混在する場合でも、複数種類のデータ間の相関の算出をすることができる相関分析装置及びその方法を提供することにある。

課題を解決するための手段

[0008] 上記課題を解決するために、本発明は、イベント毎に生起されるデータであって、種類または時間解像度の異なる複数のデータを、前記複数のデータ間の関連性を管理する関連性情報を用いて、各データ共通のデータ形式のデータに変換し、前記変換された各データの相関を算出することを特徴とする。

発明の効果

[0009] 本発明によれば、複数種類のデータが混在する場合でも、複数種類のデータ間の相関の算出をすることができる。

図面の簡単な説明

[0010] [図1]本発明に係る計算機システムの第1実施例を示す構成図である。

[図2]センサ系原データの構成図である。

[図3]テキスト系原データの構成図である。

[図4]センサ系特徴量データの構成図である。

[図5]テキスト系特徴量データの構成図である。

[図6]多重時間スケール数値特徴量データの構成図である。

[図7]特徴量・キーワード対応テーブルの構成図である。

[図8]特徴量・キーワード変換係数の構成図である。

[図9]センサ系特徴量抽出処理を説明するためのフローチャートである。

[図10]テキスト系特徴量抽出処理を説明するためのフローチャートである。

[図11]相関算出処理を説明するためのフローチャートである。

[図12]本発明に係る計算機システムの第2実施例を示す構成図である。

[図13]重要度データの構成図である。

[図14]重要度算出・部分相関算出処理を説明するためのフローチャートである。

[図15]近似相関算出処理を説明するためのフローチャートである。

発明を実施するための形態

[0011] 以下、本発明の実施例を図面に基づいて説明する。

(第1実施例)

本実施例では、複数種類のデータ間の相関を算出可能とする例を説明する。

[0012] 図1は、本発明に係る計算機システムの第1実施例を示す構成図である。

図1において、計算機システムは、相関分析サーバ1001と、複数の端末1002と、ネットワーク1003とから構成され、相関分析サーバ1001と複数の端末1002が、それぞれネットワーク1003に接続される。相関分析サーバ1001は、相関を分析するための相関分析装置である。各端末1002は、ユーザに結果を提示したり、ユーザの操作入力を受け付けたりする端末である。なお、端末1002は、2つ以上から構成されていて、各ユーザが異なる分析結果の提示や要求を入力しても良く、既存のサーバクライアント方式などの構成によって、複数の端末1002から個別の操作を行うことが可能である。

[0013] ネットワーク1003は、相関分析サーバ1001および端末1002と情報通信を行うネットワークである。ネットワーク1003は、後述する操作指示や結果の表示を伝達できる通信が可能な通信方式であれば、どのようなものでもかまわない。また、その他の構成例として、ネットワーク1003を介さずに、端末1002を相関分析サーバ1001に直接接続する構成とすることも可能である。入力データの例として、センサデータ1004お

よびテキストデータ１００５を想定している。

[0014] 次に、相関分析サーバ１００１内部の構成を説明する。相関分析サーバ１００１は、当該発明に関連する構成要素として、各種分析処理を行う処理部１０１０と、各種データを蓄積する記憶装置１０２０と、データ収集インタフェース（Ｉ／Ｆ）１０３０と、表示インタフェース（Ｉ／Ｆ）１０４０とから構成される。

[0015] 処理部１０１０は、センサデータ１００４またはテキストデータ１００５を原データ１０２１として記憶装置１０２０に格納し、格納した原データ１０２１を管理するデータ管理部１０１１と、後述する手順で原データ１０２１から特徴量を抽出する特徴量抽出部１０１２と、特徴量から相関を算出する相関算出部１０１３と、相関算出部１０１３で算出された相関を各種表示によってユーザに提示するための表示内容を構成する表示内容構成部１０１４を有するＣＰＵ（Central Processing Unit）で構成される。

[0016] 記憶装置１０２０には、センサデータやテキストデータ等の原データ１０２１、特徴量データ１０２２、相関データ１０２３などの各種データが保管される。

[0017] データ収集Ｉ／Ｆ１０３０は、センサデータ１００４およびテキストデータ１００５などのデータを収集して、処理部１０１０中の、データ管理部１０１１に入力するためのインタフェースである。

[0018] 表示Ｉ／Ｆ１０４０は、処理部１０１０中の、表示内容構成部１０１４の表示内容を、実際に端末１００２に表示するためのインタフェースである。

[0019] 図２は、センサデータ等のセンサ系原データの構成図である。図２において、センサ系原データ２０１０は、設備に設置された計測機器（センサ）や運転制御盤の情報を抽出したデータ等からなるセンサ系の原データ（計測データ）であって、データＩＤ２０１１と、記述内容２０１２から構成される。

[0020] データＩＤ２０１１は、原データ１０２１のうち、センサデータ１００４から得られた原データ２０１０が、どの設備のどのセンサで発生したかを識

別するIDである。また、記述内容2012は、データの発生した日時を示す日時2013と、データ発生元のセンサや計測装置を示すセンサID2014と、センサの出力値を示すセンサ出力値2015と、センサによるイベントの生起を示すフラグ2016から構成される。フラグ2016には、センサによるイベントの生起がある場合、「Y（あり）」が記録され、センサによるイベントの生起がない場合、「N（なし）」が記録される。

[0021] なお、多くのセンサでは出力値またはフラグの値は一つであるが、複合的に計測する装置を一まとめにしてひとつのセンサとみなす場合もあり、一般に一つ以上の出力値とフラグの値を持つことができるとする。

[0022] ここで、イベントとは、各種センサが計測データを収集することや作業者が業務記録等を作成することを意味する。また、イベントで生起されるデータとは、各種センサ（計測機器）の計測による計測データと、作業者の作成による業務記録等をデジタルデータに変換した言語データ、例えば、テキストデータを含むデータである。

[0023] 図3は、テキストデータ等のテキスト系原データの構成図である。図3において、テキスト系原データ3010は、日報や各種報告情報等、作業員によるテキスト入力を起源とするテキスト系の原データであって、データID3011と、記述内容3012から構成される。データID3011は、原データ1021のうち、テキストデータ1005から得られた原データ3011が、どの設備や業務で発生したかを識別するIDである。記述内容3012は、データの発生した日時を示す日時3013と、報告者を識別するための報告書ID3014と、自然言語にて、作業内容、結果、特記事項等の報告事項を記入するための報告事項3015から構成される。

[0024] 図4は、センサ系特徴量データの構成図である。図4において、センサ系特徴量データ4010は、センサ系原データ2010を変換して、特徴量とした統合特徴量データであって、種別ID4011と、データID4012と、起点時刻4013と、キーワード4014と、数値特徴量4015と、パラメータ4016から構成される。種別ID4011は、原データ201

0がテキスト系であるか、センサ系であるかの種別をあらわす指標である。なお、原データ2010が、センサ系の場合、それがどのような種類のセンサなのか等、さらに詳細な種別を示してもよい。データID4012は、このデータがどの設備や業務で発生したかを識別するIDである。起点時刻4013は、データの発生時刻を示す情報である。キーワード4014は、過去の特徴量パターンとキーワードの対応事例から、特徴量に対して典型的なキーワードを推定し、推定したキーワードを付与したものである。その推定方法は後述する。

[0025] 数値特徴量4015は、センサ系原データ2010を、定めた形式の特徴量に変換し、変換した特徴量を数値化したものである。例えば、計測データを、計測データの最大値よりも大きい設定値で割り算し、この割り算で得られた値を1以下の数値特徴量データとして求めたものである。この際、定めた形式とは、異なった時間スケール（1日と1時間等）での相関算出を可能とするため、特徴量を複数の時刻スケールに展開して多重のスケールで表現したものである。計測データから得られた数値特徴量データを複数の時刻スケールに展開した場合、各時刻スケールの数値特徴量データは、時間解像度の異なる時系列データとして管理される。

[0026] 例えば、数値特徴量1～100までが1時間単位の数値特徴量を示す時系列データとして管理され、数値特徴量101～200までが1日単位の数値特徴量を示す時系列データとして管理される。なお、時系列データの中に不足するデータが存在する場合には、同一の時系列データまたは他の時系列データを基に補間演算や平均値演算などを実行することで、不足分を充足することができる。なお、特徴量部分の時間スケール展開の内容については、図6で別途説明する。

[0027] パラメータ4016は、計測データに付加される情報であって、異なった時間遅れでの相関算出を決めるために用いられる情報である。例えば、パラメータ1は、種別ID4011がAで、データID4012がXの特徴量と、本特徴量Xとの間の相関の時間遅れ（相関の対象となるデータが得られた

時間と、本特微量Xが得られた時間との差）が2時間であることを示す。この際、相関の対象となるデータ（相関の相手となるデータ）が得られた時間と、本特微量Xが得られた時間との差が2時間である場合には、時間解像度の条件を満たし、両者のデータの間には、マッチングが成立することを意味する。

[0028] パラメータMは、種別ID4011がZ、データID4012がYの特微量と、本特微量Yとの時間遅れが12時間であることを示す。これらのパラメータの値は、種別ID4011やデータID4012ごとに特徴的な値を参照して決定される。ここで、一つの原因データ単独では、特微量の時間スケールをカバーできない場合には、複数の時刻での同じIDを持つセンサでの原因データを集めて一つの特微量とする。

[0029] 図5は、テキスト系特微量データの構成図である。図5において、テキスト系特微量データ5010は、テキスト系原因データ3010を変換して、特微量とした統合特微量データであって、種別ID5011と、データID5012と、起点時刻5013と、キーワード5014と、数値特微量5015と、パラメータ5016から構成される。種別ID5011は、原因データ3010がテキスト系であるか、センサ系であるかの種別をあらわす指標である。原因データ3010がテキスト系の場合、日報なのか、不具合報告書なのか等、さらに詳細な種別を示してもよい。データID5012は、このデータがどの設備や業務で発生したかを識別するIDである。起点時刻5013は、データの発生時刻を示す情報である。キーワード5014は、原因データ3010のテキストデータから、文書処理によってキーワードを抽出したものである。この際、あらかじめ決まった語彙中に存在するもののみ残してもよい。数値特微量5015は、テキストデータから対応する特微量を推定し、推定した特微量を数値化したものである。

[0030] その代表的な方法は、過去のセンサ系のデータ値パターンとキーワードの対応事例から、キーワードに対して特微量を付与するものである。その手順は、過去に同時に生起したとわかっている、テキストデータ中のキーワード

とセンサ系データの値のパターンを対応付け、特定のキーワードに対して生じやすい特徴量を対応表にして保持する。その対応表、例えば、特徴量・キーワード対応テーブルを用いて、今回抽出された各キーワードに対して、それぞれ最も典型的な特徴量を与える。最後に各抽出キーワードに対して与えられた特徴量の平均値をとることで、今回のキーワードに対する推定特徴量とする。

[0031] 上述した特徴量推定の方法は、過去のデータに基づくものであったが、その他の方法として、文書データ中の時間表現等の明らかな特徴に係る表現を活用して数値特徴量に反映する方法も可能である。例えば、テキストデータの情報として、「午後から次第に異音」という報告があったとすると、文書解析によって、キーワードを抽出し、「午後から」、「次第に」、「異音」を特徴量に反映する。

[0032] 反映の方法は、音に関連するセンサであれば、時間別に並んだ特徴量の関連するフィールドの午後からの値を与え、「次第に」という表現から、時間が経過するほどその値を大きくする、等である。

[0033] パラメータ5016は、図4と同様に、異なった時間遅れでの相関算出を決めるために用いる。記述内容は、図4と同じである。

[0034] 図6は、多重時間スケール数値特徴量データの構成図である。図6において、多重時間スケール数値特徴量データ6010は、図4の数値特徴量4015の具体的な配列例を示したものである。なお、図6では、説明を簡単にする為、時間スケール（時間解像度）が3種類の場合が示されている。時間スケール1のフィールドは、例えば、1時間ごとにセンサの値を時間順に並べたものである。時間スケール2のフィールドは、例えば、午前の測定値と午後の測定値を入力したものである。スケール3のフィールドは、例えば、1日1回の測定値を入力したものである。

[0035] 一般に、原特徴量では、センサなどの種類や、測定ルールによって、これらのスケールの全てが測定されることは稀である。そのため、欠損しているフィールドは、他のフィールドから推定して補間を行う。例えば、最も短い

時間スケールでの測定値が与えられている場合には、それらの平均値、最頻値、中央値、特定時刻での値等、なんらかの代表値を算出して他のフィールドの推定値とすればよい。逆に、より短い時間スケールの測定値を推定する場合、スケールの大きい測定値と全てのフィールドに同じ値を入力する方法で算出する方法が最も単純な方法である。測定値が原理的に長いスケールでのみ可能なものは、そのように補間するしか方法がない。あるいは、代表的な時刻のみ上位のスケールの値を入力して、他のフィールドは、何らかの欠損値を示す記号か、ゼロ値など、センシングの種類に応じてふさわしい規定値を入れることも可能である。

[0036] 一方、原理的には、現在の原データにあるよりもさらに短い時間間隔で測定が可能であり、同様のセンサによる類似のデータの測定値が参照できる場合には、スケールの大きい測定値をその変化パターンと同じ形になるように、重み付け配分することが可能である。以上のいずれかの補間方法によって、変換された後の特徴量では、全てのスケールにおいて、何らかの値もしくは、欠損を示す記号が入力されているものとする。

[0037] 以上、図6では、センサ値の種類が1種類の場合の例を示した。図2で例示したように、複数種類のデータを扱う場合には、それぞれの種類ごとに、多重の時間スケールをもつように変換する。その際のフィールドの順序は、種類ごとに並べたものをスケールごとに並べるか、一つのセンサについてスケールごとに並べたデータを種類ごとに並べるか、どちらか一つに定めて並べるものとする。

[0038] 図7は、特徴量・キーワード対応テーブルの構成図である。図7において、特徴量・キーワード対応テーブル7010は、数値特徴量からキーワードの推定や、逆にキーワードから数値特徴量の推定に用いるテーブルであって、ID7011と、代表特徴量7012と、キーワード／重み7013から構成される。ID7011と、代表特徴量7012と、キーワード／重み7013は、種類または時間解像度の異なる複数のデータを、複数のデータ間の関連性を管理するための関連性情報として構成される。ID7011は、

特徴量を識別するための識別子である。代表特徴量 7012 は、特徴量のうち代表となる特徴量を数値化したものである。キーワード／重み 7013 は、特徴量に対応するキーワードと、キーワードの重みから構成され、上段が、キーワードを示し、下段が、その重みを示す。

[0039] この際、ID7011 と、代表特徴量 7012 およびキーワード／重み 7013 は、数値特徴量 4015 とキーワード 4014 との対応関係が定義された関連性情報（第 1 の関連性情報）と、数値特徴量 5015 とキーワード 5014 との対応関係が定義された関連性情報（第 2 の関連性情報）として構成され、ID7011 には、イベント毎に生起されるデータを特定する情報として、場所、時間、設備または作業者を特定する情報が格納される。

[0040] 特徴量・キーワード対応テーブル 7010 を構成するに際しては、まず、なるべく多くの対応するキーワードと特徴量が共存するデータを準備する。それらのデータは人手で準備しても良いし、過去に明確に共起した特徴量とテキストのデータを付き合わせて生成しても良い。そのような複数のデータ対をもって、そのまま特徴量・キーワード対応テーブル 7010 とすることも可能であるが、対応のばらつきを回避するため、以下の処理を行う。

[0041] まず、各データの特徴量の「数値的近さ」を基にグループ分けを行う。ここで、数値的近さとして、例えば、特徴量の並びを一つの数値ベクトルとみなしてユークリッド距離を用いる。近さを一つ決めると、グループ分けを行う処理は、クラスタリングと呼ばれる。クラスタリングを行った後、各グループに属する特徴量から代表特徴量を算出する。算出の方法は、例えば、特徴量の並びを一つの数値ベクトルとみなしてそれらの平均ベクトルを算出するなどの方法が可能である。

[0042] 一方、各クラスタを構成する特徴量が、それぞれ対として持っていたキーワードをクラスタ内のデータで全て集めて各キーワードの頻度を集計する。大きい頻度を持つキーワードから順に定めておいた個数のキーワードをキーワードテーブルの対応キーワードとして採用する。同時に、頻度そのもの、

または、キーワード全体で、総和が1.0になるように、頻度を規格化したものを、キーワードの重みとして用いる。以上の処理で、クラスごとに、代表特徴量（ベクトル）と重み付きのキーワード群の対が算出される。なお、各対応データのIDは、対応データを定めるごとに任意に定める。

[0043] キーワードと特徴量を関連付けるその他の方法として、上記クラスタリングを用いずに、特徴量ベクトルと各単語の頻度を表わすベクトルの対応付けを直接算出する方法を用いてもよい。その手順は、上記と同様の共存データ中で、上記クラスタリング時に用いた方法と同様に、特徴量を数値ベクトルとみなし、一方で、同じ共存データ中で各キーワードの出現頻度をベクトル表示したものを準備する（以後、キーワード出現頻度ベクトルと呼ぶ）。以上の結果、一つの共存データ中で、特徴量ベクトルとキーワードの出現頻度ベクトルの対が出来るので、それらの関係を互いに変換する際に、最も誤差が少なく変換できるように、変換関数のパラメータを決定すればよい。

[0044] 変換関数の表現としては、例えば線形変換を仮定する方法が代表的である。この係数は、一つの係数行列と一つの係数ベクトルで表わされる。その係数行列をA、係数ベクトルをbと置くと、特徴量ベクトルからキーワード出現頻度ベクトルを予測する式は、

$$(\text{キーワード出現頻度ベクトル}) = A \times (\text{特徴量ベクトル}) + b$$

と表わされる。ここで、行列Aと特徴量ベクトルとの掛け算は、通常の行列とベクトルの積であり、ベクトルbの足し算は、ベクトルの間の足し算である。

[0045] 逆に、キーワード出現頻度ベクトルから、特徴量ベクトルを予測する式は、（特徴量ベクトル）と（キーワード出現頻度ベクトル）を入れ替えて、上記の逆関数を決定すればよい。なお、以上の係数の算出手順は、多変量解析における重回帰分析の手法を用いればよい。

[0046] 図8に、その線形変換を仮定した変換関数の例を示す。図8において、特徴量・キーワード変換係数4201は、上記係数行列Aであり、特徴量・キーワード変換係数4202は、係数ベクトルbである。すなわち、係数行列

Aの各成分は、特徴量ベクトルに対するキーワード出現頻度ベクトルの関連の大きさを示しており、係数ベクトルbは、特徴量・キーワード対応テーブル4201で変換された後のキーワードベクトルの大きさを適正化するために足される値である。

[0047] 次に、センサ系特徴量抽出処理を図9のフローチャートに従って説明する。この処理は、センサ系原データを、特徴量・キーワード対応テーブル7010を用いて、センサ系特徴量に変換する手順を示したものである。

[0048] まず、特徴量抽出部1012は、記憶装置102に格納された原データ1021に対して、どのような種類のデータであるのかを判別し、原データ1020が、テキスト系原データ2010であると判別した場合、原データ2010にデータID4012を割り当てる(S901)。次に、特徴量抽出部1012は、原データ2010を基にデータの種別に応じて特徴量の整形を行う(S902)。特徴量の整形とは、欠値の補間や、データの種別に対して、標準的な時刻での計測に合致しない場合などに、同じくデータ補間で時間をずらした値に変換する処理等である。

[0049] 次に、特徴量抽出部1012は、整形された特徴量に対して、特徴量の多重スケール化を行う(S903)。方法は、図6で説明したとおりである。次に、特徴量抽出部1012は、図4でも述べたように、キーワードの推定を行う(S904)。キーワード推定の方法は、過去の特徴量パターンとキーワードの対応事例から特徴量に対して典型的なキーワードを付与するものである。

[0050] その方法の一例として、図7で上述した、特徴量・キーワードテーブル7010を用いる方法を説明する。まず、与えられた特徴量パターンに対して、特徴量・キーワードテーブル7010中の代表特徴量7012との近さが、最も近い代表特徴量を選出する。ここで、特徴量間の近さは、例えば、特徴量・キーワードテーブル7010を作成する際に用いたユークリッド距離を用いて算出する。

[0051] 最も近い代表特徴量を選んだら、それに対応する特徴量・キーワードテー

ブル 7 0 1 0 中のキーワードを推定し、推定したキーワードを推定キーワードとする。

[0052] 最後に、特徴量抽出部 1 0 1 2 は、特徴量のパラメータを決定する（S 9 0 5）。パラメータは、各特徴量を総当り的に対にして、各対の相関を算出する際に、それぞれどのような時間遅れで相関を算出すべきであることを示すものである。

[0053] これらのパラメータは、特徴量の組み合わせに対する特性から導かれるものであり、基本は人手で算出する。例えば、あるデータ ID 4 0 1 2 の特徴量が、ある設備の容器内圧力を示すものであって、別のデータ ID 4 0 1 2 の特徴量は、その設備に付属した装置の温度を示すものであったとして、それらは 1 時間の遅れで高い相関が出ることが知られている場合、相関を算出すべきデータ間の時間遅れを 1 時間とする。あるいは、過去のデータを複数の時間遅れを用いて分析することで、相関の高い時間遅れを経験的にパラメータとしてもよい。

[0054] 次に、テキスト系特徴量抽出処理を図 1 0 のフローチャートに従って説明する。この処理は、テキスト系原データを、特徴量・キーワード対応テーブル 7 0 1 0 を用いて、テキスト系特徴量に変換する手順を示したものである。

[0055] 特徴量抽出部 1 0 1 2 は、記憶装置 1 0 2 0 に格納された原データ 1 0 2 1 に対して、どのような種類のデータであるのかを判別し、原データ 1 0 2 1 が、テキスト系原データ 3 0 1 0 であると判別した場合、原データ 3 0 1 0 にデータ ID 5 0 1 2 を割り当てる（S 1 0 0 1）。

[0056] 次に、特徴量抽出部 1 0 1 2 は、テキストデータ 1 0 0 5 から得られた原データ 3 0 1 0 について、単語分割や品詞の同定などの自然言語処理の技法を用いてテキスト処理を行う（S 1 0 0 2）。処理された結果は、一般にキーワードとして使いづらい一般的過ぎる単語、数詞、自立語でない助詞などを含む。このため、特徴量抽出部 1 0 1 2 は、頻度、品詞、その他の特徴を用いてキーワードとなりうるものを原データ 3 0 1 0 中から抽出する（S 1

003)。

[0057] 次に、図9で説明した手順とは逆に、特徴量抽出部1012は、特徴量・キーワードテーブル7010を用いて、キーワードから換算特徴量を抽出する(S1004)。その手順は、例えば、特徴量・キーワードテーブル7010の中で、現在のデータで抽出されたキーワードを含む行を全て抽出し、キーワードの重みによって、抽出された行を加重平均したものを換算特徴量とする。

[0058] 例えば、キーワードの「蒸気圧」が、特徴量・キーワードテーブル7010の中の1, 2行目に、キーワードの「高温」が、特徴量・キーワードテーブル7010の中の1, 3行目にあった場合、1行目の特徴量に対する重みは、「蒸気圧」と「高温」の由来の重みを加えたものとし、2行目の特徴量に対する重みは、「蒸気圧」の重みとし、3行目は、同様に「高温」の重みとして、3つの代表特徴量間で加重平均を行う。

[0059] 最後に、特徴量抽出部1012は、パラメータの決定を行う(S1005)。パラメータの決定は、図9の処理と同様に、データの内容によって、人手で与えたものを用いるか、過去のデータを用いて決定する。

[0060] 次に、相関算出処理を図11のフローチャートに従って説明する。この処理は、図9と図10の処理で変換された2つの特徴量間の相関を算出する手順を示したものである。マッチングの判定では、双方の特徴量のパラメータをみて、合致すればマッチングしたとして、相関の算出を行い、それ以外は、双方の特徴量のパラメータが合致しないとして、不整合を示す値を返すこととしている。

[0061] まず、相関算出部1013は、記憶装置1020に格納された特徴量データ1022の中から、対象となる2つのデータの特徴量を入力する(S1101)。次に、相関算出部1013は、入力された特徴量から各種IDを読み取る(S1102)。ここで、各種IDとは、前述の、特徴量の種類を示す種別IDや、データの出所を示すIDである。

[0062] 次に、相関算出部1013は、相手特徴量の各種IDに対応したパラメー

タを読み取る（S 1 1 0 3）。前述したように、パラメータは、マッチングが可能な相手に対する遅れ時間が記述されている。

[0063] 次に、相関算出部 1 0 1 3 は、読み取った各パラメータはマッチングするか否かを判定し（S 1 1 0 4）、パラメータを検討した結果、マッチングが可能な組み合わせであれば、そのパラメータによって指定される時間遅れを伴って、特徴量の相関値を算出する（S 1 1 0 5）。例えば、相関算出部 1 0 1 3 は、読み取った各パラメータの時間遅れが、時間解像度の条件を満たし、両者の特徴量の間に、マッチングが成立した場合、両者の特徴量の相関値を算出する。

[0064] 続いて、相関算出部 1 0 1 3 は、キーワード間の共起度合いも算出する（S 1 1 0 6）。例えば、相関算出部 1 0 1 3 は、キーワード 4 0 1 4 とキーワード 5 0 1 4 の中から、読み取った各パラメータの時間遅れが、時間解像度の条件を満たす、キーワード 4 0 1 4 とキーワード 5 0 1 4 を選択し、選択したキーワード 4 0 1 4 とキーワード 5 0 1 4 間の共起度合いを算出する。

[0065] ここで、キーワードの共起度合いとは、共通したキーワードが、互いの特徴量の間で何割あるかを示す量である。

例えば、
$$\frac{\text{（共通キーワード数）}}{\{ \text{（データ 1 の全キーワード数）} \times \text{（データ 2 の全キーワード数）} \}}$$
という式を用いて算出する。

[0066] 一方、パラメータを検討した結果、マッチング不可と判定した場合、相関算出部 1 0 1 3 は、「不整合」の判定結果を返す（S 1 1 0 7）。ここで、時間遅れの量が、互いの特徴量に格納されている時間幅よりも大きい場合も、マッチング不可の判断となる。

[0067] 本実施例では、特徴量抽出部 1 0 1 2 は、イベント毎に生起されるデータであって、種類または時間解像度の異なる複数のデータを、複数のデータ間の関連性を管理する特徴量・キーワード対応テーブル 7 0 1 0 を用いて、各データ共通のデータ形式のデータに変換するデータ変換部として機能し、相関算出部 1 0 1 3 は、特徴量抽出部 1 0 1 2 で得られた各データの相関を算

出する相関算出部として機能する。

[0068] この際、特徴量抽出部 1012 は、イベント毎に生起されるデータが、計測機器の計測による計測データまたは作業者の業務記録を示す言語データ（テキストデータ）であって、イベント毎に生起されるデータの発生時間を示す時間解像度が異なる複数の時系列データで構成されている場合、計測データの中から、当該計測データの特徴量を示す数値特徴量（第 1 の数値特徴量データ）4015 を時間解像度毎に抽出し、各抽出した数値特徴量 4015 を特徴量・キーワード対応テーブル（関連性情報）7010 を基に、言語データに関連するキーワード（第 1 の言語データ）4014 に変換すると共に、言語データの中から、イベントに関連するキーワード（第 2 の言語データ）5014 を時間解像度毎に抽出し、各抽出したキーワード 5014 を数値特徴量（第 1 の数値特徴量データ）5015 を基に、キーワード 5014 の特徴量を示す数値特徴量（第 2 の数値特徴量データ）5015 に変換するデータ変換部として機能する。

[0069] また、相関算出部 1013 は、特徴量抽出部 1012 で得られた数値特徴量 4015 と数値特徴量 5015 の中から、計測データまたは言語データ（テキストデータ）に付加されるパラメータ 4016、5016 で特定される時間解像度の条件を満たす、数値特徴量 4015 と数値特徴量 5015（マッチングが成立する数値特徴量）を選択し、選択した数値特徴量 4015 と数値特徴量 5015 との相関を算出すると共に、特徴量抽出部 1012 で得られたキーワード 4014 とキーワード 5014 の中から、計測データまたは言語データ（テキストデータ）に付加されるパラメータ 4016、5016 で特定される時間解像度の条件を満たす、キーワード 4014 とキーワード 5014（マッチングが成立するキーワード）を選択し、選択したキーワード 4014 とキーワード 5014 との相関を算出する相関算出部として機能する。

[0070] また、特徴量抽出部 1012 は、各時間解像度（時間スケール）の数値特徴量データの中に不足する数値特徴量データが存在する場合、不足する数値

特徴量データを、当該数値特徴量データとは時間解像度が異なる他の時間解像度の数値特徴量データまたは不足する数値特徴量データと同一の時間解像度の数値特徴量データから推定するデータ変換部として機能する。

[0071] 本実施例によれば、特徴量・キーワード対応テーブル7010を用いることで、複数種類のデータとして、テキストデータ（言語データ）と計測データが混在した場合でも、複数種類のデータ間の相関の算出をすることができる。

[0072] また、本実施例によれば、異なった種別のデータ間の相関を適切な時間スケール、例えば、時間遅れで算出可能となる。

[0073] また、本実施例によれば、重要度・相関規準を設定して、部分データの相関のみ算出することで、広範囲のデータの関連性を監視できるにもかかわらず、計算量やデータ量を抑制することが可能である。また、重要度が高い関係から順にデータを残しつつ、データ量を必要に応じて変更することが可能となる。

[0074] （第2実施例）

本実施例では、第1実施例で述べた相関の算出方法に加えて、重要度・相関基準によって、部分データのみ相関を算出する処理格納方法についての実施例を説明する。

[0075] 第1実施例の方法により、異なった種類のデータ間であっても相関を算出することが可能であるが、全ての組み合わせを算出することは現実のシステム運用においては、計算時間及び、記憶装置の容量の問題から困難である。そのため、何らかの方法で部分的な算出を元に近似的な関係性を提示する方法が必要となる。本実施例では、そのため、データの重要度を算出することで、その方法を提供するものである。

[0076] 図12は、本発明に係る計算機システムの第2実施例を示す構成図である。本実施例は、処理部1010中に、相関算出部1013で算出された相関を基に、相関が算出されていない別の要素間の相関を算出する近似相関算出部1015が追加され、記憶装置1020中に、重要度データ1024が追

加されたものであり、他の構成は、図１に示す第１の実施例と同様である。

[0077] 図１３は、本実施例で用いる重要度データの構成図である。図１３において、重要度データ１０２４は、データＩＤ１３０１と、原重要度１３０２と、伝播重要度１３０３と、伝播元１３０４から構成され、各行１３１１～１３１４が１つのデータに相当する。なお、算出の手順は後述する。

[0078] データＩＤ１３０１は、重要度データ１０２４を識別するための識別子である。原重要度１３０２は、各データの本来の重要度を示す情報である。伝播重要度１３０３は、相関の状況と原重要度によって算出される情報である。伝播元１３０４は、データの伝播元を示す情報である。

[0079] 次に、重要度算出・部分相関算出処理を図１４のフローチャートに従って説明する。この処理は、本実施例における、上述した重要度の算出と、部分データのみの相関算出を行う手順である。本手順は、センサデータ１００４またはテキストデータ１００５が、新規データとして一つ加えられる度に、その新規データと全ての既存データに対して行われる手順である。

[0080] 相関算出部１０１３は、まず、入力された新規データの特徴量を入力する（Ｓ１４０１）、次に、入力された新規データに対して、原重要度を決定する（Ｓ１４０２）。新規データの原重要度は、過去の原因分析上、主要要因となった頻度や、重要なイベントに関連していたか等を勘案して決定される。

[0081] 次に、相関算出部１０１３は、既存データを順に読み込み（Ｓ１４０３）、読み込んだ各既存データと新規データとの相関算出手順に進む。その相関算出手順では、まず、相関算出部１０１３は、既に算出されている既存データの伝播重要度（定義は後述）を読み込み（Ｓ１４０９）、既存データと新規データとの相関を算出する（Ｓ１４０５）。

[0082] 次に、相関算出部１０１３は、算出した相関の値から、連結重要度という指標を算出する（Ｓ１４０６）。ここで、連結重要度とは、データの対に対して、一つ決まる指標であり、相関と重要度の双方の大きさを一つの数値で表わすものである。すなわち、連結重要度は、相関が大きく、重要度が大き

いほど大きな値をとる量である。

[0083] その最も単純な定義は、

$$(\text{連結重要度}) = (\text{既存データの伝播重要度}) \times (\text{相関の大きさ})$$

とすればよい。

[0084] 次に、相関算出部 1013 は、連結重要度が保存閾値よりも大きいかな否かを判定し (S1407)、この連結重要度が、あらかじめ定めた保存閾値より大きい場合には、相関を保存し (S1408)、新規データの伝播重要度を算出する (S1409)。

[0085] 新規データの伝播重要度は、

例えば、

$$(\text{伝播重要度}) = (\text{原重要度}) + (\text{係数}) \times (\text{既存データの伝播重要度}) \times (\text{相関の大きさ})$$

で定義される。ここで、係数は、あらかじめ定められた値であり、過去の全データの平均結合数の逆数など、伝播重要度が過度に大きな値とならないための係数であり、通常は、1.0でよい。

[0086] 次に、相関算出部 1013 は、既存データの伝播重要度を更新する (S1410)。

[0087] 更新の方法は、

例えば、

$$(\text{更新後伝播重要度}) = (\text{更新前伝播重要度}) + (\text{係数}) \times (\text{新規データの伝播重要度}) \times (\text{相関の大きさ})$$

で定義される。ここで、係数は、あらかじめ定められた値であり、上記の新規データの伝播重要度の場合と同様に設定される。

[0088] 一方、ステップ S1407 にて、連結重要度があらかじめ定めた保存閾値より小さいと判定した場合、あるいは、ステップ S1409 の処理の後、相関算出部 1013 は、既存データの処理は全て終了したかな否かを判定し (S1411)、次の既存データの処理が存在する場合には、ステップ S1409 の処理に戻り、次の既存データに対して、ステップ S1409～ステップ S1411 の処理を繰り返し、既存データの処理が全て終了したと判定した

場合には、このルーチンでの処理を繰り返す。

[0089] 次に、近似相関算出処理を図15のフローチャートに従って説明する。この処理は、任意のデータ間の相関を近似的に算出する手順を示すものである。

[0090] まず、近似相関算出部1015は、相関を算出したいデータ対を指定する(S1501)。近似相関算出部1015は、指定されたデータ対に対して、上述した保存相関が存在するかを判断し(S1502)、保存相関が存在する場合は、その保存された相関を出力する(S1503)。

[0091] 一方、ステップS1502で、保存相関が存在しないと判定した場合、近似相関算出部1015は、各データの重要度を算出し、両重要度が共に算出閾値より大きいと判定し(S1504)、両重要度が共に算出閾値より大きいと判定した場合には、両データの相関を算出し(S1505)、算出閾値より小さい重要度のデータが存在し、両重要度が共に算出閾値より大きくないと判定した場合には、両データの相関を近似的にゼロとする(S1506)。ここで、重要度は、そのデータに伝播重要度が存在する場合には、その値を、伝播重要度が存在しない場合には、原重要度を用いる。

[0092] 本実施例によれば、イベント個別の重要度と相関の大きさを用いて、イベント間の連結の重要度を算出し、算出結果を基に部分データの相関のみを算出するようにしたので、計算量とデータ量を削減することができる。

[0093] 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。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、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることも可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

[0094] 例えば、特徴量抽出部1012は、種類の異なる複数のデータが、計測データと言語データである場合、計測データの中から、当該計測データの特徴

量を示す数値特徴量（第１の数値特徴量データ）４０１４を抽出し、抽出した数値特徴量４０１４を特徴量・キーワード対応テーブル７０１０を基に、キーワード（第１の言語データ）４０１４に変換すると共に、言語データの中から、イベントに関連するキーワード（第２の言語データ）５０１４を抽出し、抽出したキーワード５０１４を特徴量・キーワード対応テーブル７０１０を基に、数値特徴量（第２の数値特徴量データ）５０１５に変換するデータ変換部として機能する。

[0095] 相関算出部１０１３は、特徴量抽出部１０１２で得られた数値特徴量４０１４と数値特徴量５０１５を基に数値特徴量４０１４と数値特徴量５０１５との相関を算出し、特徴量抽出部１０１２で得られたキーワード４０１４とキーワード５０１４とを基にキーワード４０１４とキーワード５０１４との相関を算出する相関算出部として機能することができる。

[0096] 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、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリや、ハードディスク、ＳＳＤ（Solid State Drive）等の記録装置、または、ＩＣ（Integrated Circuit）カード、ＳＤ（Secure Digital）メモリカード、ＤＶＤ（Digital Versatile Disc）等の記録媒体に記録して置くことができる。

符号の説明

[0097] １００１ 相関分析サーバ、１００２ 端末、１００４ センサデータ、１００５ テキストデータ、１０１０ 処理部、１０２０ 記憶装置、１０１１ データ管理部、１０１２ 特徴量抽出部、１０１３ 相関算出部、１０１４ 表示内容構成部。

請求の範囲

[請求項1] 複数のイベント間の関連性を分析し、分析対象のイベントに関連する他のイベントを抽出する相関分析装置において、

前記イベント毎に生起されるデータであって、種類または時間解像度の異なる複数のデータを、前記複数のデータ間の関連性を管理する関連性情報を用いて、各データ共通のデータ形式のデータに変換するデータ変換部と、

前記データ変換部で変換された各データの相関を算出する相関算出部と、を有することを特徴とする相関分析装置。

[請求項2] 請求項1に記載の相関分析装置において、

前記データ変換部は、

前記種類の異なる複数のデータが、計測機器の計測による計測データと、作業者の業務記録を示す言語データである場合、前記計測データの中から、当該計測データの特徴量を示す第1の数値特徴量データを抽出し、前記抽出した第1の数値特徴量データを前記関連性情報を基に、前記言語データに関連する第1の言語データに変換すると共に、前記言語データの中から、前記イベントに関連する第2の言語データを抽出し、前記抽出した第2の言語データを前記関連性情報を基に、当該第2の言語データの特徴量を示す第2の数値特徴量データに変換し、

前記相関算出部は、

前記データ変換部で抽出された第1の数値特徴量データと前記データ変換部で変換された第2の数値特徴量データとを基に前記第1の数値特徴量データと前記第2の数値特徴量データとの相関を算出すると共に、前記データ変換部で変換された第1の言語データと前記データ変換部で抽出された第2の言語データとを基に前記第1の言語データと前記第2の言語データとの相関を算出することを特徴とする相関分析装置。

- [請求項3] 請求項2に記載の相関分析装置において、
前記関連性情報は、
前記第1の数値特徴量データと前記第1の言語データとの対応関係が定義された第1の関連性情報と、前記第2の数値特徴量データと前記第2の言語データとの対応関係が定義された第2の関連性情報とから構成され、前記第1の関連性情報と前記第2の関連性情報は、前記イベント毎に生起されるデータを特定する情報であって、場所、時間、設備または作業者を特定する情報を含むことを特徴とする相関分析装置。
- [請求項4] 請求項1に記載の相関分析装置において、
前記データ変換部は、
前記時間解像度の異なる複数のデータが、前記イベント毎に生起されるデータの発生時間を示す時間解像度が異なる複数の時系列データで構成されている場合、前記各時間解像度の時系列データを数値特徴量データに変換し、
前記相関算出部は、
前記データ変換部で変換された各数値特徴量データの相関を算出することを特徴とする相関分析装置。
- [請求項5] 請求項4に記載の相関分析装置において、
前記データ変換部は、
前記変換された前記各時間解像度の数値特徴量データの中に不足する数値特徴量データが存在する場合、前記不足する数値特徴量データを、当該数値特徴量データとは時間解像度が異なる他の時間解像度の数値特徴量データまたは前記不足する数値特徴量データと同一の時間解像度の数値特徴量データから推定することを特徴とする相関分析装置。
- [請求項6] 請求項1に記載の相関分析装置において、
前記データ変換部は、

前記イベント毎に生起されるデータが、計測機器の計測による計測データまたは作業者の業務記録を示す言語データであって、前記イベント毎に生起されるデータの発生時間を示す時間解像度が異なる複数の時系列データで構成されている場合、前記計測データの中から、当該計測データの特徴量を示す第1の数値特徴量データを前記時間解像度毎に抽出し、前記各抽出した第1の数値特徴量データを前記関連性情報を基に、前記言語データに関連する第1の言語データに変換すると共に、前記言語データの中から、前記イベントに関連する第2の言語データを前記時間解像度毎に抽出し、前記各抽出した第2の言語データを前記関連性情報を基に、当該第2の言語データの特徴量を示す第2の数値特徴量データに変換し、

前記相関算出部は、

前記データ変換部で抽出された第1の数値特徴量データと前記データ変換部で変換された第2の数値特徴量データの中から、前記計測データまたは前記言語データに付加されるパラメータで特定される時間解像度の条件を満たす数値特徴量データを選択し、前記選択した第1の数値特徴量データと第2の数値特徴量データとの相関を算出すると共に、前記データ変換部で変換された第1の言語データと前記データ変換部で抽出された第2の言語データの中から、前記計測データまたは前記言語データに付加されるパラメータで特定される時間解像度の条件を満たす言語データを選択し、前記選択した第1の言語データと第2の言語データとの相関を算出することを特徴とする相関分析装置。

[請求項7]

請求項1に記載の相関分析装置であって、

前記相関算出部は、

前記データ変換部で変換された各データの中に新規データと既存データが存在する場合、前記既存データ個別の重要度及び相関の大きさを基に前記既存データの連結の重要度を算出し、前記算出結果が閾値

を超えた場合、前記新規データと前記既存データとの相関を算出することを特徴とする相関分析装置。

[請求項8] 請求項7に記載の相関分析装置であって、
 前記相関算出部は、

 前記既存データ個別の重要度を、前記新規データ個別の重要度及び相関の大きさを基に更新することを特徴とする相関分析装置。

[請求項9] 複数のイベント間の関連性を分析し、分析対象のイベントに関連する他のイベントを抽出する相関分析方法において、

 前記イベント毎に生起されるデータであって、種類または時間解像度の異なる複数のデータを、前記複数のデータ間の関連性を管理する関連性情報を用いて、各データ共通のデータ形式のデータに変換するデータ変換ステップと、

 前記データ変換ステップで変換された各データの相関を算出する相関算出ステップと、を含むことを特徴とする相関分析方法。

[請求項10] 請求項9に記載の相関分析方法において、

 前記データ変換ステップでは、

 前記種類の異なる複数のデータが、計測機器の計測による計測データと、作業者の業務記録を示す言語データである場合、前記計測データの中から、当該計測データの特徴量を示す第1の数値特徴量データを抽出し、前記抽出した第1の数値特徴量データを前記関連性情報を基に、前記言語データに関連する第1の言語データに変換すると共に、前記言語データの中から、前記イベントに関連する第2の言語データを抽出し、前記抽出した第2の言語データを前記関連性情報を基に、当該第2の言語データの特徴量を示す第2の数値特徴量データに変換し、

 前記相関算出ステップでは、

 前記データ変換ステップで抽出された第1の数値特徴量データと前記データ変換ステップで変換された第2の数値特徴量データとを基に

前記第 1 の数値特徴量データと前記第 2 の数値特徴量データとの相関を算出すると共に、前記データ変換ステップで変換された第 1 の言語データと前記データ変換ステップで抽出された第 2 の言語データとを基に前記第 1 の言語データと前記第 2 の言語データとの相関を算出することを特徴とする相関分析方法。

[請求項11]

請求項 10 に記載の相関分析方法において、
前記関連性情報は、

前記第 1 の数値特徴量データと前記第 1 の言語データとの対応関係が定義された第 1 の関連性情報と、前記第 2 の数値特徴量データと前記第 2 の言語データとの対応関係が定義された第 2 の関連性情報とから構成され、前記第 1 の関連性情報と前記第 2 の関連性情報は、前記イベント毎に生起されるデータを特定する情報であって、場所、時間、設備または作業者を特定する情報を含むことを特徴とする相関分析方法。

[請求項12]

請求項 9 に記載の相関分析方法において、
前記データ変換ステップでは、

前記時間解像度の異なる複数のデータが、前記イベント毎に生起されるデータの発生時間を示す時間解像度が異なる複数の時系列データで構成されている場合、前記各時間解像度の時系列データを数値特徴量データに変換し、

前記相関算出ステップでは、

前記データ変換ステップで変換された各数値特徴量データの相関を算出することを特徴とする相関分析方法。

[請求項13]

請求項 12 に記載の相関分析方法において、
前記データ変換では、

前記変換された前記各時間解像度の数値特徴量データの中に不足する数値特徴量データが存在する場合、前記不足する数値特徴量データを、当該数値特徴量データとは時間解像度が異なる他の時間解像度の

数値特徴量データまたは前記不足する数値特徴量データと同一の時間解像度の数値特徴量データから推定することを特徴とする相関分析方法。

[請求項14]

請求項9に記載の相関分析方法において、

前記データ変換ステップでは、

前記イベント毎に生起されるデータが、計測機器の計測による計測データまたは作業者の業務記録を示す言語データであって、前記イベント毎に生起されるデータの発生時間を示す時間解像度が異なる複数の時系列データで構成されている場合、前記計測データの中から、当該計測データの特徴量を示す第1の数値特徴量データを前記時間解像度毎に抽出し、前記各抽出した第1の数値特徴量データを前記関連性情報を基に、前記言語データに関連する第1の言語データに変換すると共に、前記言語データの中から、前記イベントに関連する第2の言語データを前記時間解像度毎に抽出し、前記各抽出した第2の言語データを前記関連性情報を基に、当該第2の言語データの特徴量を示す第2の数値特徴量データに変換し、

前記相関算出ステップでは、

前記データ変換ステップで抽出された第1の数値特徴量データと前記データ変換ステップで変換された第2の数値特徴量データの中から、前記計測データまたは前記言語データに付加されるパラメータで特定される時間解像度の条件を満たす数値特徴量データを選択し、前記選択した第1の数値特徴量データと第2の数値特徴量データとの相関を算出すると共に、前記データ変換ステップで変換された第1の言語データと前記データ変換ステップで抽出された第2の言語データの中から、前記計測データまたは前記言語データに付加されるパラメータで特定される時間解像度の条件を満たす言語データを選択し、前記選択した第1の言語データと第2の言語データとの相関を算出することを特徴とする相関分析方法。

[請求項15]

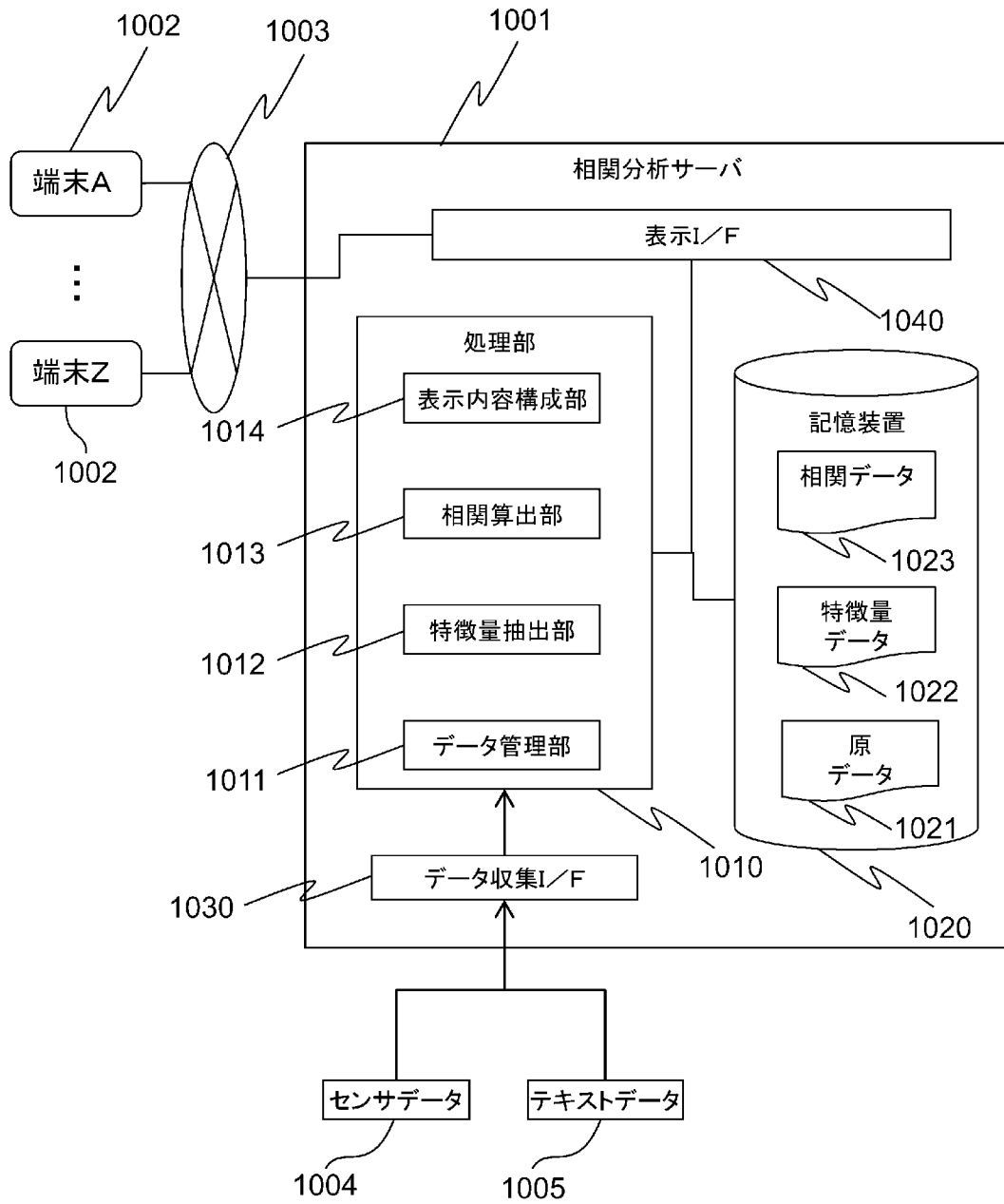
請求項 9 に記載の相関分析方法であって、

前記相関算出ステップでは、

前記データ変換ステップで変換された各データの中に新規データと既存データが存在する場合、前記既存データ個別の重要度及び相関の大きさを基に前記既存データの連結の重要度を算出し、前記算出結果が閾値を超えた場合、前記新規データと前記既存データとの相関を算出することを特徴とする相関分析方法。

[図1]

図 1



[図2]

図 2

データID	記述内容							
V5060 7080	日時	センサID	〇〇値	...	△△値	□□ フラグ	...	〇〇 フラグ
	2011/12/ 12 18:30	A123- 9876	60.5 °C	...	10kPa	Y	...	N

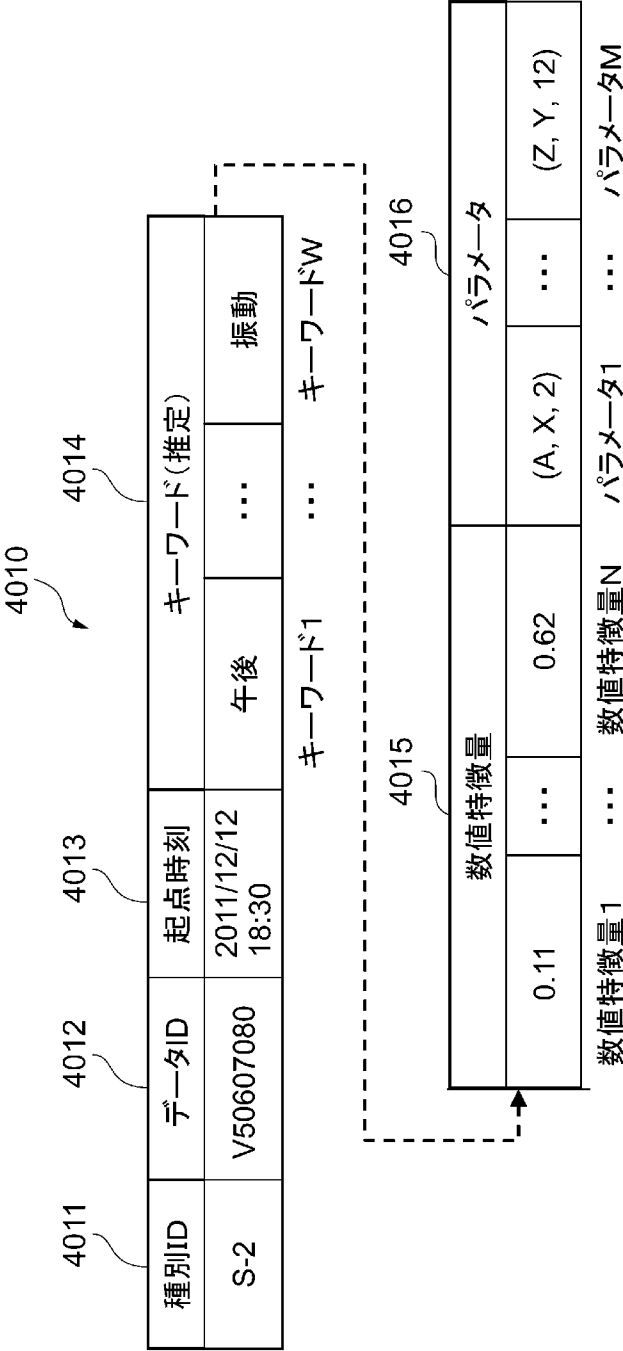
[図3]

図 3

データID	記述内容		
	日時	報告者ID	報告事項
T12345678	2011/12/12 12:15	1000001	〇〇検査 △△異常

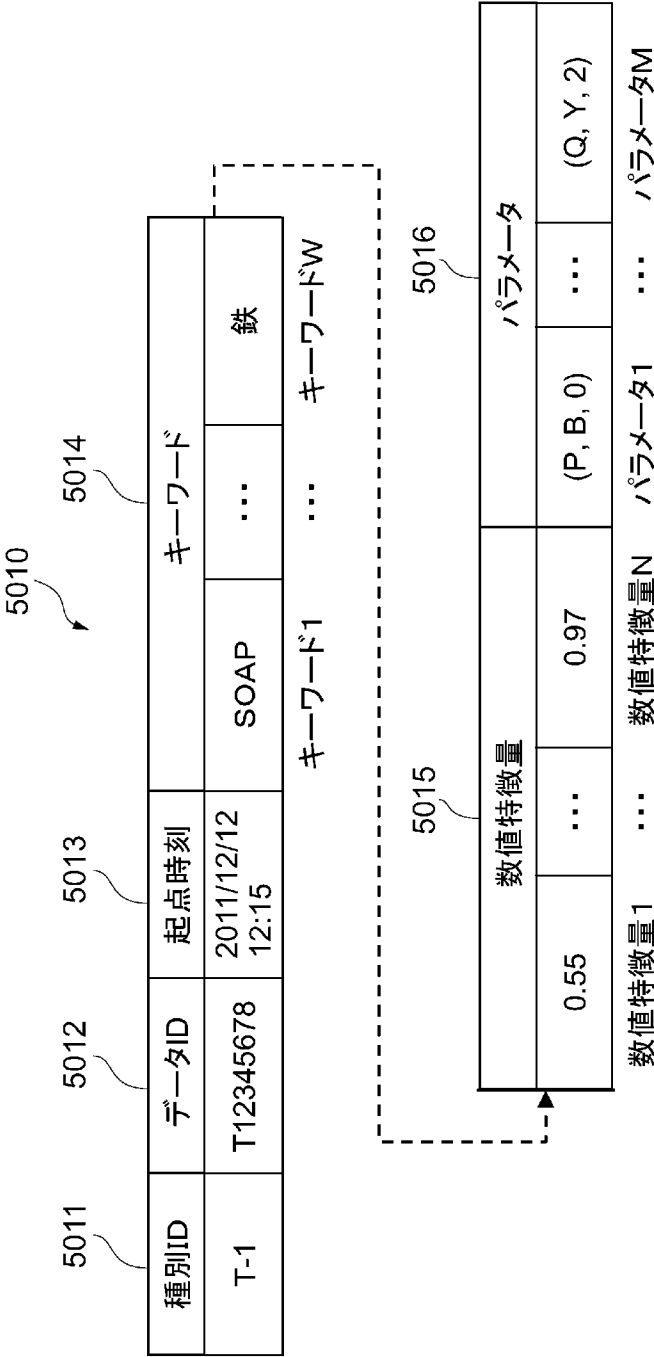
[図4]

図 4



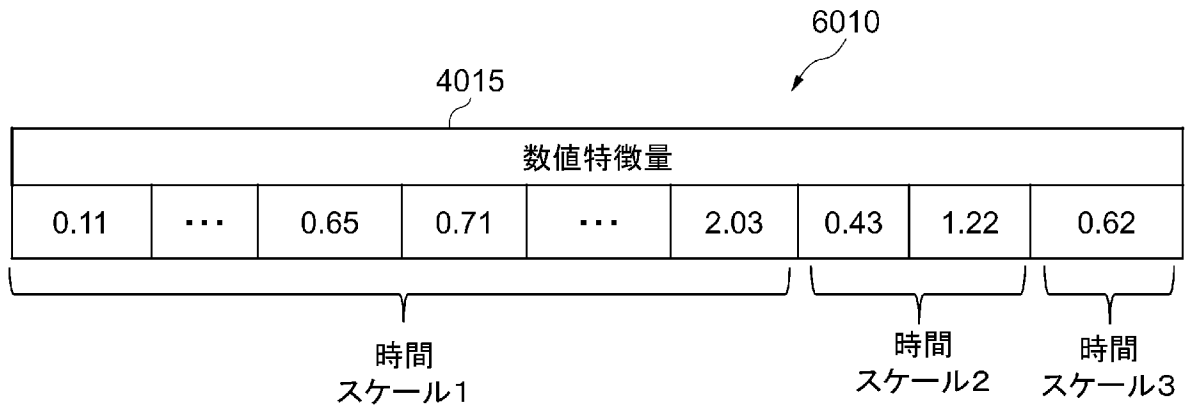
[図5]

図 5



[図6]

図 6



[図7]

図 7

Figure 7 shows a table with three main columns: ID (7011), 代表特徴量 (7012), and キーワード／重み (7013). The table is labeled 7010.

ID	代表特徴量			キーワード／重み		
0001	0.10	...	0.46	高温	...	蒸気圧
				0.25	...	0.01
...	...					
9876	0.03	...	0.79	振動	...	午前
				0.33	...	0.02

[図8]

図 8

	特徴量1	...	特徴量N
キーワード1	0.2	...	0.5
⋮	⋮		⋮
キーワードM	0.3	...	1.0

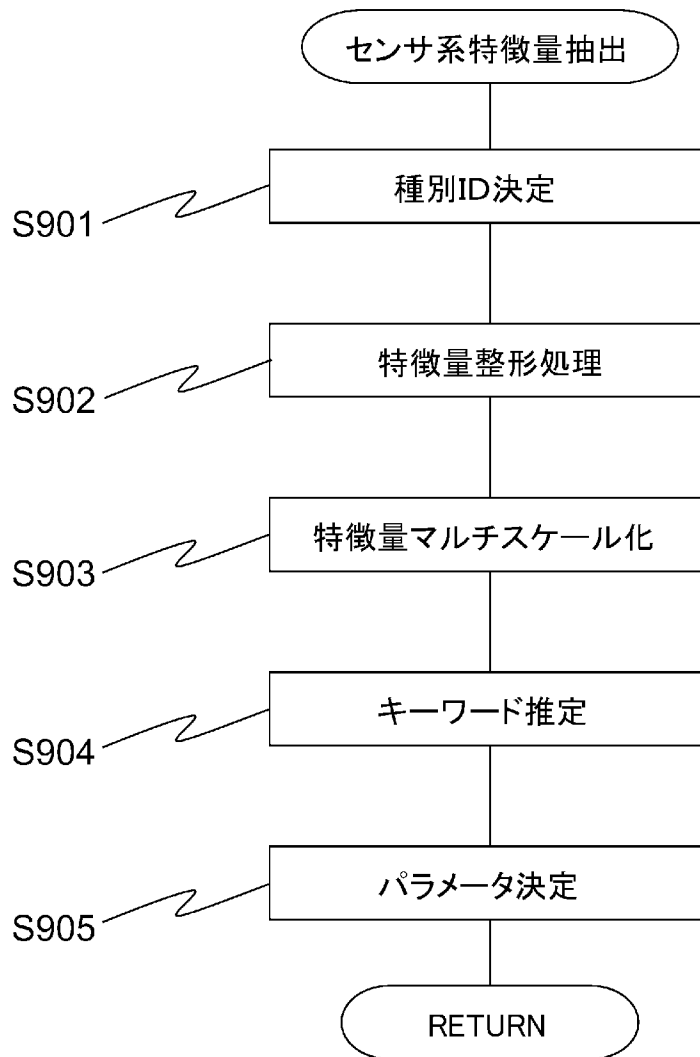
4201

	加算係数
キーワード1	1.0
⋮	⋮
キーワードM	2.5

4202

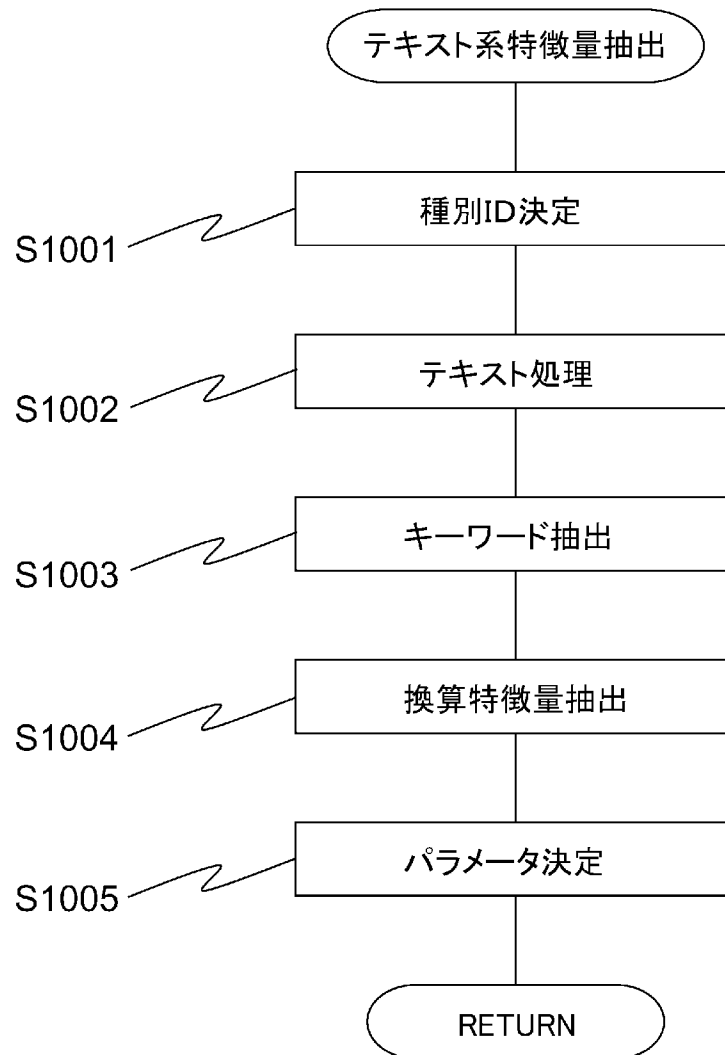
[図9]

図 9



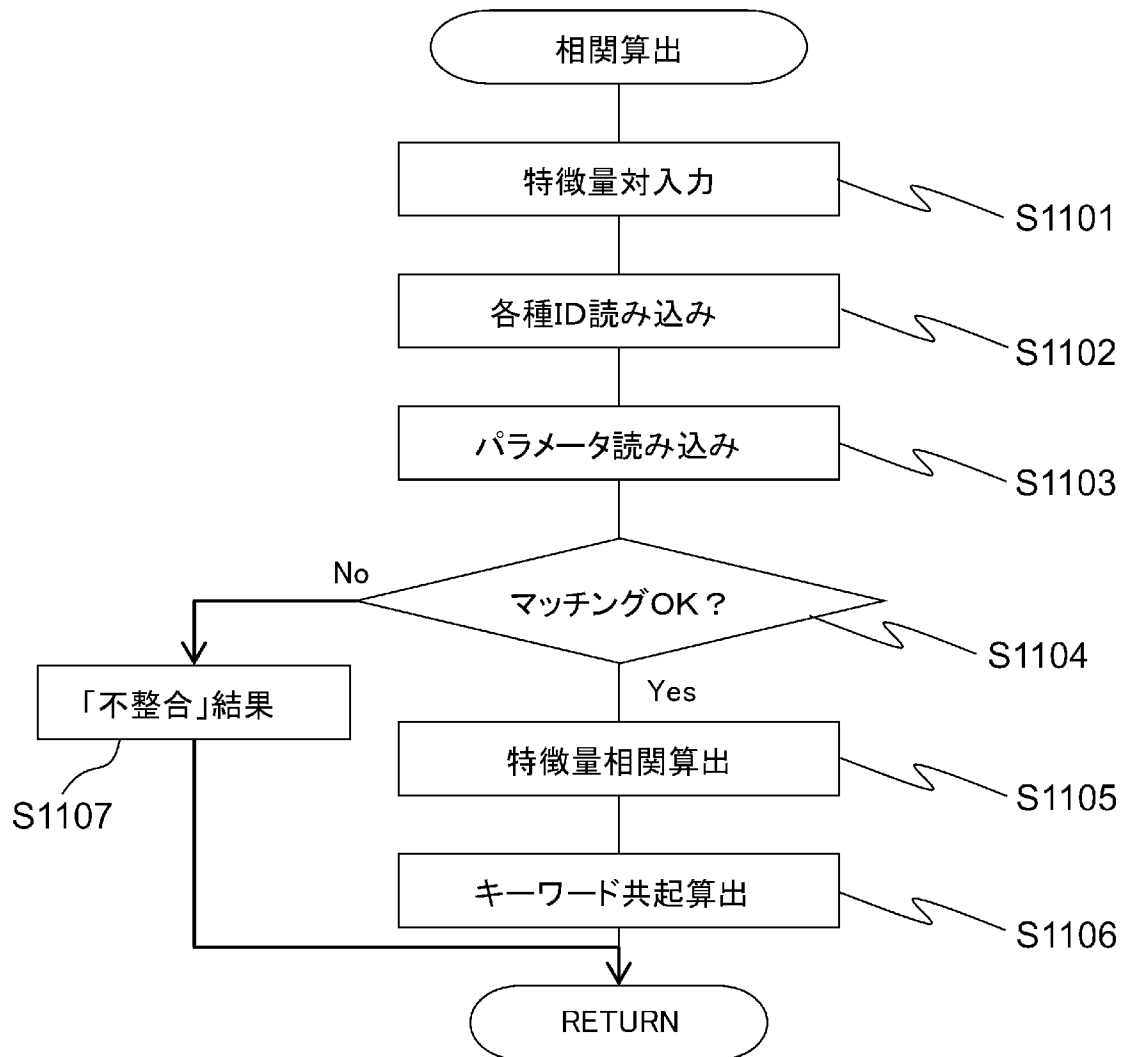
[図10]

図 1 0



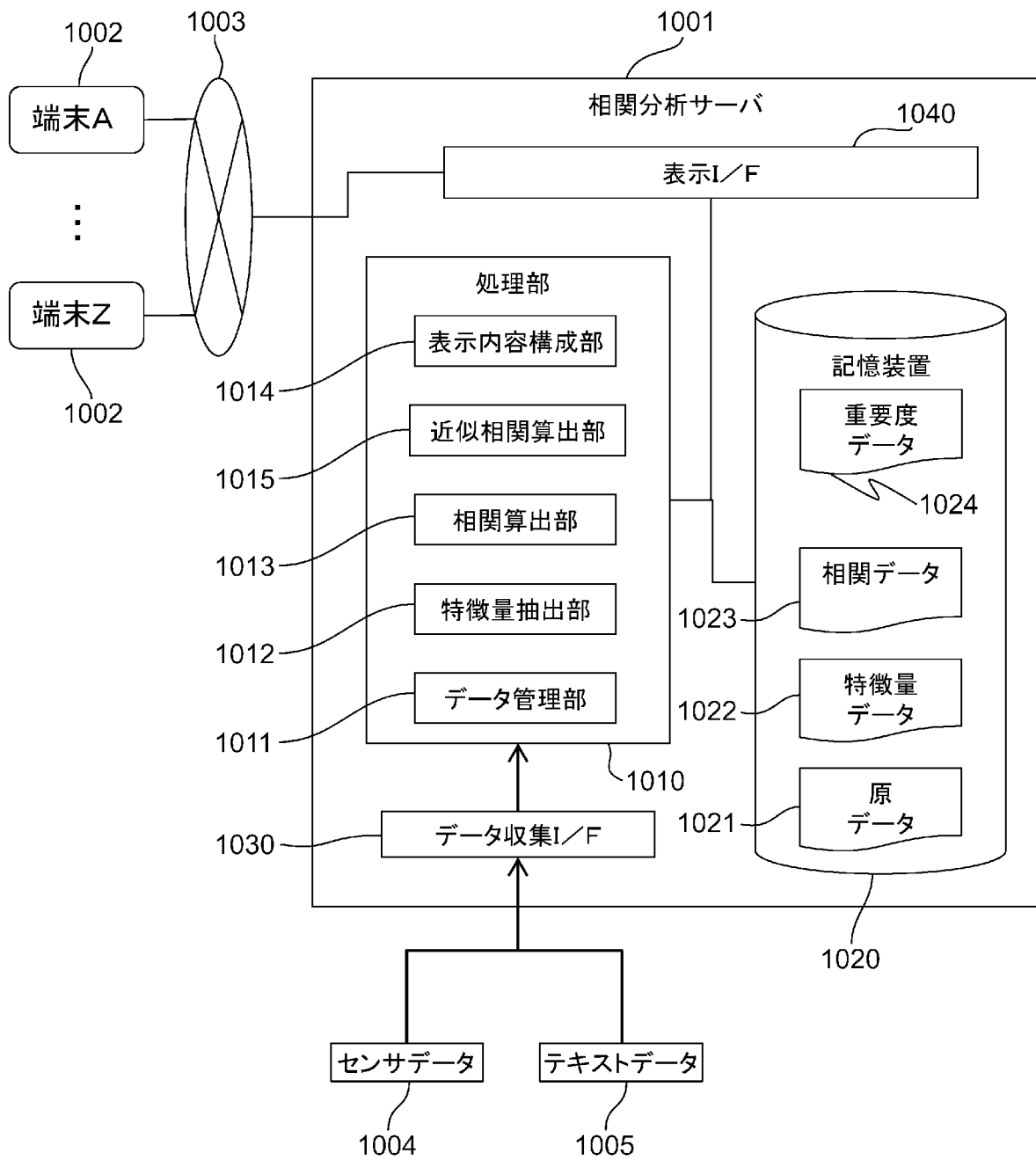
[図11]

図 1 1



[図12]

図 1 2

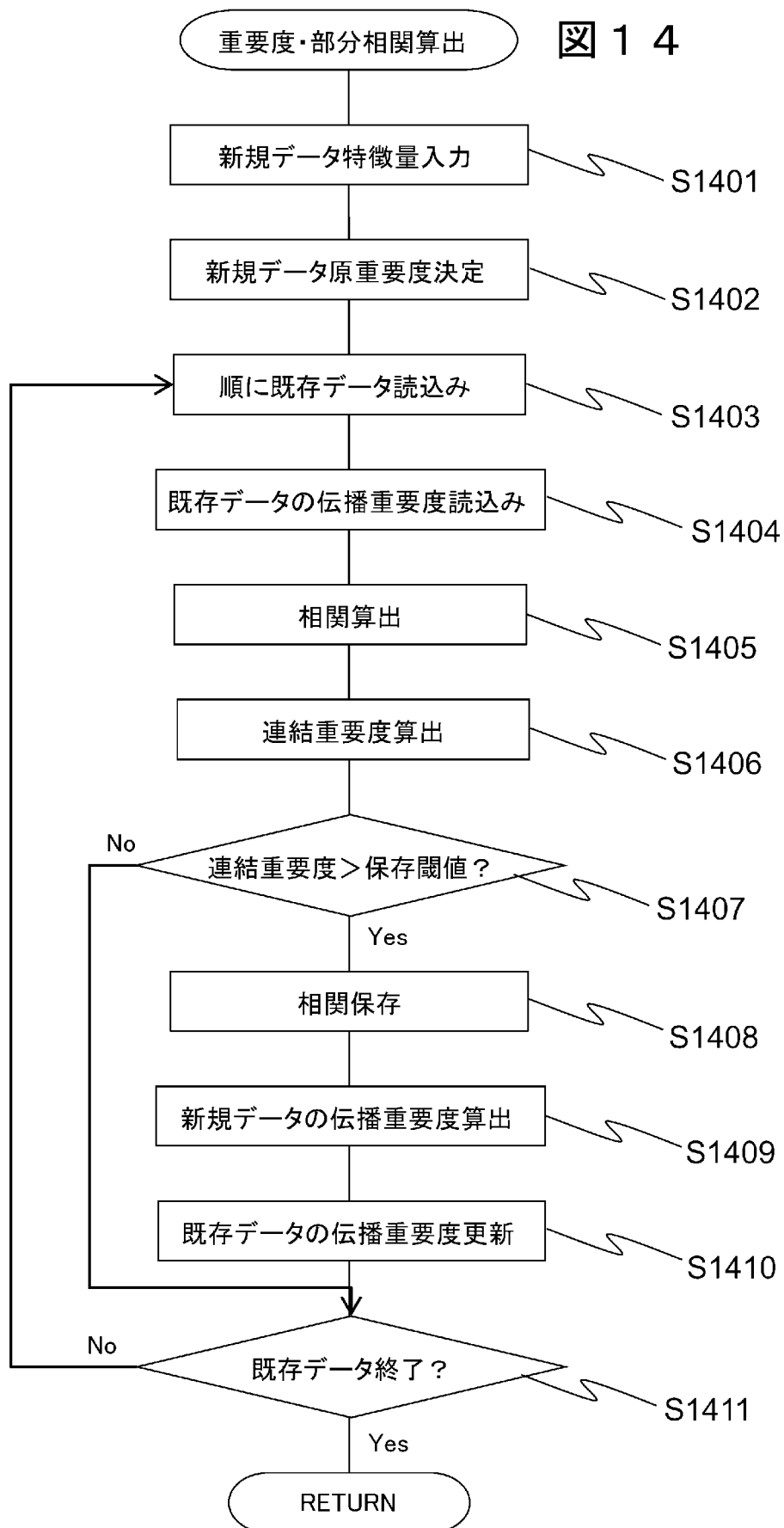


[図13]

図 1 3

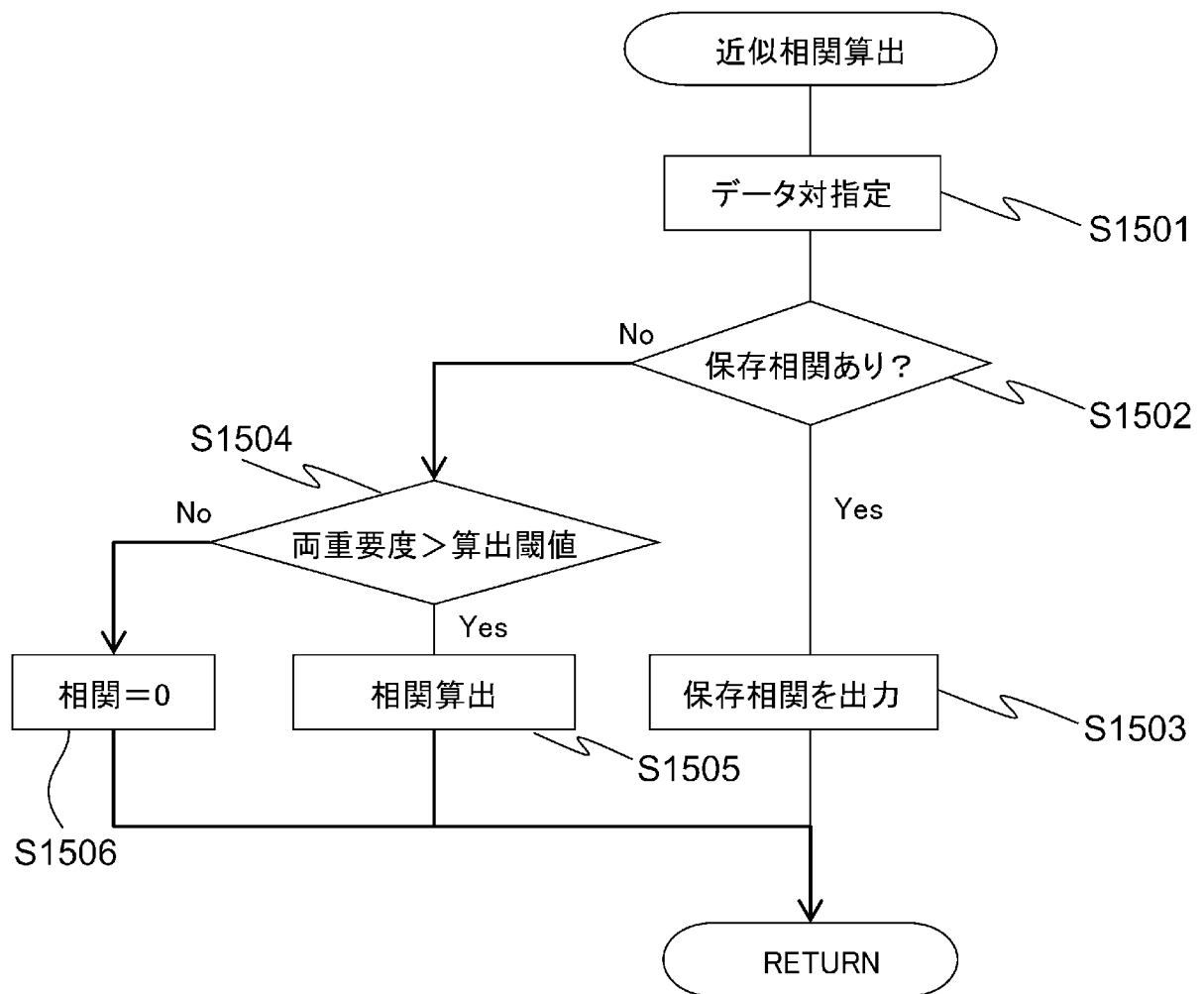
	1301	1302	1024 1303	1304
	データID	原重要度	伝播重要度	伝播元
1311	T12345678	6.0	8.0	T12345679
1312	T12345679	0.0	0.0	—
1313	T12345680	0.0	1.0	T12345678, T12345690
	...			
1314	V50607081	3.0	3.0	—

[図14]



[図15]

図 1 5



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/065552

A. CLASSIFICATION OF SUBJECT MATTER

G06Q50/04 (2012.01) i

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)

G06Q50/04

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

Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2012

Kokai Jitsuyo Shinan Koho 1971-2012 Toroku Jitsuyo Shinan Koho 1994-2012

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 Y A	JP 11-25169 A (Toshiba Corp.), 29 January 1999 (29.01.1999), columns 22 to 109; fig. 1 to 20 & US 6230064 B1	1, 7-9, 15 4-5, 12-13 2-3, 6, 10-11, 14
Y	JP 2002-15000 A (Toshiba Corp.), 18 January 2002 (18.01.2002), columns 1 to 49; fig. 1 (Family: none)	4-5, 12-13
Y	JP 2006-31378 A (Nippon Hoso Kyokai), 02 February 2006 (02.02.2006), columns 1 to 30 (Family: none)	5, 13

☒ 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
19 July, 2012 (19.07.12)Date of mailing of the international search report
31 July, 2012 (31.07.12)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/065552

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 2008-234618 A (Oki Electric Industry Co., Ltd.), 02 October 2008 (02.10.2008), columns 40 to 113 (Family: none)	7-8, 15
A	JP 2009-265905 A (Toshiba Corp.), 12 November 2009 (12.11.2009), entire text; all drawings (Family: none)	1-15
A	JP 8-95959 A (Hitachi, Ltd.), 07 February 1996 (07.02.1996), entire text; all drawings (Family: none)	1-15
A	JP 2007-279840 A (Omron Corp.), 25 October 2007 (25.10.2007), entire text; all drawings & US 2007/0239651 A1 & EP 1860512 A2 & KR 10-2007-0099442 A & CN 101052294 A & TW 200817861 A	1-15

A. 発明の属する分野の分類 (国際特許分類 (I P C))

Int.Cl. G06Q50/04(2012.01)i

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 (I P C))

Int.Cl. G06Q50/04

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 2 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 2 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 2 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 Y A	JP 11-25169 A (株式会社東芝) 1999.01.29, 第 22-109 欄, 第 1-20 図 & US 6230064 B1	1, 7-9, 15 4-5, 12-13 2-3, 6, 10-11, 14
Y	JP 2002-15000 A (株式会社東芝) 2002.01.18, 第 1-49 欄, 第 1 図 (ファミリーなし)	4-5, 12-13
Y	JP 2006-31378 A (日本放送協会) 2006.02.02, 第 1-30 欄 (ファミ	5, 13

☒ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、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

1 9 . 0 7 . 2 0 1 2

国際調査報告の発送日

3 1 . 0 7 . 2 0 1 2

国際調査機関の名称及びあて先

日本国特許庁 (I S A / J P)

郵便番号 1 0 0 - 8 9 1 5

東京都千代田区霞が関三丁目 4 番 3 号

特許庁審査官 (権限のある職員)

田付 徳雄

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 6 2

5 L

3 2 4 3

C (続き) . 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
	リーなし)	
A	JP 2008-234618 A (沖電気工業株式会社) 2008.10.02, 第 40-113 欄 (ファミリーなし)	7-8, 15
A	JP 2009-265905 A (株式会社東芝) 2009.11.12, 全文, 全図 (ファミリーなし)	1-15
A	JP 8-95959 A (株式会社日立製作所) 1996.02.07, 全文, 全図 (ファミリーなし)	1-15
A	JP 2007-279840 A (オムロン株式会社) 2007.10.25, 全文, 全図 & US 2007/0239651 A1 & EP 1860512 A2 & KR 10-2007-0099442 A & CN 101052294 A & TW 200817861 A	1-15