(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4879178号
(P4879178)
(45)発行日　平成24年2月22日(2012.2.22)　(24)登録日　平成23年12月9日(2011.12.9)

(51)Int.Cl.　F I
G06T 7/00 (2006.01)　G06T 7/00 350Z

請求項の数 28 (全 20 頁)

(21)出願番号　特願2007-529054(P2007-529054)
(86)(22)出願日　平成17年8月1日(2005.8.1)
(65)公表番号　特表2008-508645(P2008-508645A)
(43)公表日　平成20年3月21日(2008.3.21)
(86)国際出願番号　PCT/IB2005/052570
(87)国際公開番号　WO2006/013549
(87)国際公開日　平成18年2月9日(2006.2.9)
審査請求日　平成20年7月22日(2008.7.22)
(31)優先権主張番号　60/592,911
(32)優先日　平成16年8月2日(2004.8.2)
(33)優先権主張国　米国(US)

(73)特許権者　504194580
石川　博
埼玉県富士見市ふじみ野東１丁目２２－２－１０２
(72)発明者　石川　博
愛知県東海市加木屋町北鹿持３４－９３

審査官　新井　則和

(56)参考文献　特開平０４－２７６７８５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06T 7/00

(54)【発明の名称】自動パターン解析のための方法と装置

(57)【特許請求の範囲】
【請求項１】
記憶手段と処理手段を備えた情報処理システムにより実行されるパターン解析方法であって、
少なくとも１つの第１データを受信する第１工程と、
該第１データを、データ、写像、およびパターンの記録可能な前記記憶手段上のデータ構造に記録する第２工程と、
該第１データの構造と解釈に応じて少なくとも１つの原始写像を決定する第３工程と、
該原始写像を前記データ構造に記録する第４工程と、
前記データ構造に記録された少なくとも１つの第２データを選択する第５工程と、
該第２データから少なくとも１つの第３データを誘導する複数の手続きから少なくとも１つの手続きを選択する第６工程と、
該手続きに従い前記第２データから前記第３データを誘導する第７工程と、
前記データ構造に該第３データを記録する第８工程と
を有し、前記複数の手続きに、前記データ構造に記録された少なくとも１つの第１写像を所定の方法により選択し該第１写像を前記第２データに適用することにより前記第３データを誘導することを特徴とする第１手続きと、
前記第２データ内に表現された複数の集合の直積を少なくとも１つとり該直積を前記第３データ内に表現することにより該第３データを誘導することを特徴とする第２手続きと
を含むことを特徴とするパターン解析方法であって、

前記データ構造に記録された少なくとも１つの第４データ内に所定の方法により少なくとも１つの第１パターンを探す第９工程
をさらに有することを特徴としたパターン解析方法。
【請求項２】
前記第９工程は前記第４データのエントロピーに従って前記第１パターンを探すことを特徴とした請求項１記載のパターン解析方法。
【請求項３】
前記第１パターンが見つかれば該第１パターンを前記データ構造に記録する第１０工程と、
前記データ構造に記録された少なくとも１つの第２パターンをパターン解析結果として提供する第１１工程
をさらに有することを特徴とした請求項１ないし２いずれか１項に記載のパターン解析方法。
【請求項４】
前記第１パターンが見つかれば該第１パターンに対応する理想化されたデータである少なくとも１つの第５データを所定の方法により生成し前記データ構造に記録する第１２工程
をさらに有することを特徴とした請求項１ないし３いずれか１項に記載のパターン解析方法。
【請求項５】
前記第１２工程は前記第４データ内に表現された少なくとも１つの第１確率測度を選択しエントロピーのより低い第２確率測度を該第１確率測度から生成し該第２確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第３確率測度を選択し該第３確率測度を集中させて少なくとも１つの第４確率測度を生成し該第４確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第５確率測度を選択し該第５確率測度内の少なくとも１つの確率の集中に各々対応した複数の確率測度を生成し該複数の確率測度を前記第５データ内に表現するか、
前記第４データ内の近似的に繰り返すパターンをより正確に前記第５データ内に繰り返させるか
の少なくとも１つにより前記第５データを生成することを特徴とした請求項４記載のパターン解析方法。
【請求項６】
前記第１パターンが見つかれば前記第４データに結び付けられた手続き及び写像に従ってパターン写像を決定し前記データ構造に記録する第１３工程をさらに有し、
前記第８工程は選択された前記手続きと該手続きで写像が使用されていれば該写像を前記第３データと結び付けて前記データ構造に記録することを特徴とした請求項１ないし５いずれか１項に記載のパターン解析方法。
【請求項７】
一連の工程を所定の停止条件が満たされるまで繰り返す第１４工程をさらに有し、該一連の工程は前記第５ないし１４工程のうち前記第１４工程以外の少なくとも１つを含むことを特徴とした請求項１ないし６いずれか１項に記載のパターン解析方法。
【請求項８】
前記複数の手続きに、
前記データ構造に記録された少なくとも１つの第２写像を選択し前記第２データ内に表現された少なくとも１つの第１集合の該第２写像による逆像をとり前記第３データ内に該逆像を表現することにより該第３データを誘導することを特徴とする第３手続きをさらに含むことを特徴とした請求項１ないし７いずれか１項に記載のパターン解析方法。
【請求項９】
前記複数の手続きに、

前記第２データ内に表現された少なくとも１つの第２集合の少なくとも１つの部分集合をとり前記第３データ内に該部分集合を表現することにより該第３データを誘導することを特徴とする第４手続きをさらに含むことを特徴とした請求項１ないし８いずれか１項に記載のパターン解析方法。

【請求項１０】

前記原始写像が、

恒等写像、定数写像、等号写像、積写像、複数の写像の積写像を与える写像、引戻し演算写像、射影写像、対角写像、置換写像、写像合成写像、評価写像、複数の低位写像を組み合わせて高位写像を与える写像、ＣＵＲＲＹ写像、論理演算写像、ベクトル演算写像、順序写像、汎関数演算写像、固定点演算写像の１つ以上を含むことを特徴とする請求項１ないし９いずれか１項に記載のパターン解析方法。

【請求項１１】

コンピュータープログラムを含む記憶手段と、該コンピュータープログラムを実行する時に、

少なくとも１つの第１データを受信する第１工程と、

該第１データを、データ、写像、およびパターンの記録可能な前記記憶手段上のデータ構造に記録する第２工程と、

該第１データの構造と解釈に応じて少なくとも１つの原始写像を決定する第３工程と、

該原始写像を前記データ構造に記録する第４工程と、

前記データ構造に記録された少なくとも１つの第２データを選択する第５工程と、

該第２データから少なくとも１つの第３データを誘導する複数の手続きから少なくとも１つの手続きを選択する第６工程と、

該手続きに従い前記第２データから前記第３データを誘導する第７工程と、

前記データ構造に該第３データを記録する第８工程と

を実行するように配置された処理手段を有し、前記複数の手続きに、前記データ構造に記録された少なくとも１つの第１写像を所定の方法により選択し該第１写像を前記第２データに適用することにより前記第３データを誘導することを特徴とする第１手続きと、

前記第２データ内に表現された複数の集合の直積を少なくとも１つとり該直積を前記第３データ内に表現することにより該第３データを誘導することを特徴とする第２手続きと

を含むことを特徴とするパターン解析システムであって、

前記処理手段が前記コンピュータープログラムを実行する時に、

前記データ構造に記録された少なくとも１つの第４データ内に所定の方法により少なくとも１つの第１パターンを探す第９工程

をさらに実行するように配置されたパターン解析システム。

【請求項１２】

前記処理手段が前記コンピュータープログラムを実行する時に、

前記第１パターンが見つかれば該第１パターンを前記データ構造に記録する第１０工程と、

前記データ構造に記録された少なくとも１つの第２パターンをパターン解析結果として提供する第１１工程

をさらに実行するように配置された請求項１１記載のパターン解析システム。

【請求項１３】

前記処理手段が前記コンピュータープログラムを実行する時に、

前記第１パターンが見つかれば該第１パターンに対応する理想化されたデータである少なくとも１つの第５データを所定の方法により生成し前記データ構造に記録する第１２工程

をさらに実行するように配置された請求項１１ないし１２いずれか１項に記載のパターン解析システム。

【請求項１４】

前記第１２工程は

前記第４データ内に表現された少なくとも１つの第１確率測度を選択しエントロピーのよ

り低い第２確率測度を該第１確率測度から生成し該第２確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第３確率測度を選択し該第３確率測度を集中させて少なくとも１つの第４確率測度を生成し該第４確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第５確率測度を選択し該第５確率測度内の少なくとも１つの確率の集中に各々対応した複数の確率測度を生成し該複数の確率測度を前記第５データ内に表現するか、
前記第４データ内の近似的に繰り返すパターンをより正確に前記第５データ内に繰り返させるかの少なくとも１つにより前記第５データを生成することを特徴とした請求項１３記載のパターン解析システム。
【請求項１５】
前記処理手段が前記コンピュータープログラムを実行する時に、
前記第１パターンが見つかれば前記第４データに結び付けられた手続き及び写像に従ってパターン写像を決定し前記データ構造に記録する第１３工程をさらに実行するように配置され、前記第８工程は選択された前記手続きと該手続きで写像が使用されていれば該写像を前記第３データと結び付けて前記データ構造に記録することを特徴とした請求項１１ないし１４いずれか１項に記載のパターン解析システム。
【請求項１６】
前記処理手段が前記コンピュータープログラムを実行する時に、
一連の工程を所定の停止条件が満たされるまで繰り返す第１４工程をさらに実行するように配置され、該一連の工程は前記第５ないし１４工程のうち前記第１４工程以外の少なくとも１つを含むことを特徴とした請求項１１ないし１５いずれか１項に記載のパターン解析システム。
【請求項１７】
前記複数の手続きに、
前記データ構造に記録された少なくとも１つの第２写像を選択し前記第２データ内に表現された少なくとも１つの第１集合の該第２写像による逆像をとり前記第３データ内に該逆像を表現することにより該第３データを誘導することを特徴とする第３手続きをさらに含むことを特徴とした請求項１１ないし１６いずれか１項に記載のパターン解析システム。
【請求項１８】
前記複数の手続きに、
前記第２データ内に表現された少なくとも１つの第２集合の少なくとも１つの部分集合をとり前記第３データ内に該部分集合を表現することにより該第３データを誘導することを特徴とする第４手続きをさらに含むことを特徴とした請求項１１ないし１７いずれか１項に記載のパターン解析システム。
【請求項１９】
前記原始写像が、
恒等写像、定数写像、等号写像、積写像、複数の写像の積写像を与える写像、引戻し演算写像、射影写像、対角写像、置換写像、写像合成写像、評価写像、複数の低位写像を組み合わせて高位写像を与える写像、ＣＵＲＲＹ写像、論理演算写像、ベクトル演算写像、順序写像、汎関数演算写像、固定点演算写像の１つ以上を含むことを特徴とする請求項１１ないし１８いずれか１項に記載のパターン解析システム。
【請求項２０】
記憶手段を備えた処理手段によって実行されたとき、パターン解析をするように配置されたソフトウェア・プログラムが記録されたソフトウェア記録媒体であって、該ソフトウェア・プログラムは
実行されたとき、少なくとも１つの第１データを受信する第１モジュールと、
実行されたとき、該第１データを、データ、写像、およびパターンの記録可能な前記記憶手段上のデータ構造に記録する第２モジュールと、

実行されたとき、該第１データの構造と解釈に応じて少なくとも１つの原始写像を決定する第３モジュールと、
実行されたとき、該原始写像を前記データ構造に記録する第４モジュールと、
実行されたとき、前記データ構造に記録された少なくとも１つの第２データを選択する第５モジュールと、
実行されたとき、該第２データから少なくとも１つの第３データを誘導する複数の手続きから少なくとも１つの手続きを選択する第６モジュールと、
実行されたとき、前記手続きに従い前記第２データから前記第３データを誘導する第７モジュールと、
実行されたとき、前記データ構造に該第３データを記録する第８モジュールと
を有し、前記複数の手続きに、
前記データ構造に記録された少なくとも１つの第１写像を所定の方法により選択し該第１写像を前記第２データに適用することにより前記第３データを誘導することを特徴とする第１手続きと、
前記第２データ内に表現された複数の集合の直積を少なくとも１つとり該直積を前記第３データ内に表現することにより該第３データを誘導することを特徴とする第２手続きと
を含むことを特徴とするソフトウェア記録媒体
であって、
前記ソフトウェア・プログラムが実行されたとき、前記データ構造に記録された少なくとも１つの第４データ内に所定の方法により少なくとも１つの第１パターンを探す第９モジュール
をさらに有することを特徴としたソフトウェア記録媒体。
【請求項２１】
前記ソフトウェア・プログラムが
実行されたとき、前記第１パターンが見つかれば該第１パターンを前記データ構造に記録する第１０モジュールと、
実行されたとき、前記データ構造に記録された少なくとも１つの第２パターンをパターン解析結果として提供する第１１モジュールと
をさらに有することを特徴とした請求項２０記載のソフトウェア記録媒体。
【請求項２２】
前記ソフトウェア・プログラムが
実行されたとき、前記第１パターンが見つかれば該第１パターンに対応する理想化されたデータである少なくとも１つの第５データを所定の方法により生成し前記データ構造に記録する第１２モジュール
をさらに有することを特徴とした請求項２０ないし２１いずれか１項に記載のソフトウェア記録媒体。
【請求項２３】
前記第１２モジュールは、実行されたとき、
前記第４データ内に表現された少なくとも１つの第１確率測度を選択しエントロピーのより低い第２確率測度を該第１確率測度から生成し該第２確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第３確率測度を選択し該第３確率測度を集中させて少なくとも１つの第４確率測度を生成し該第４確率測度を前記第５データ内に表現するか、
前記第４データ内に表現された少なくとも１つの第５確率測度を選択し該第５確率測度内の少なくとも１つの確率の集中に各々対応した複数の確率測度を生成し該複数の確率測度を前記第５データ内に表現するか、
前記第４データ内の近似的に繰り返すパターンをより正確に前記第５データ内に繰り返させるかの少なくとも１つにより前記第５データを生成することを特徴とした請求項２２記載のソフトウェア記録媒体。

【請求項２４】
前記ソフトウェア・プログラムが実行されたとき、前記第１パターンが見つかれば前記第４データに結び付けられた手続き及び写像に従ってパターン写像を決定し前記データ構造に記録する第１３モジュールをさらに有し、
前記第８モジュールは選択された前記手続きと該手続きで写像が使用されていれば該写像を前記第３データと結び付けて前記データ構造に記録することを特徴とした請求項２０ないし２３いずれか１項に記載のソフトウェア記録媒体。
【請求項２５】
前記ソフトウェア・プログラムが
実行されたとき、一連のモジュールを所定の停止条件が満たされるまで繰り返し実行する第１４モジュールをさらに有し、該一連のモジュールは前記第５ないし１４モジュールのうち前記第１４モジュール以外の少なくとも１つを含むことを特徴とした請求項２０ないし２４いずれか１項に記載のソフトウェア記録媒体。
【請求項２６】
前記複数の手続きに、
前記データ構造に記録された少なくとも１つの第２写像を選択し前記第２データ内に表現された少なくとも１つの第１集合の該第２写像による逆像をとり前記第３データ内に該逆像を表現することにより該第３データを誘導することを特徴とする第３手続きをさらに含むことを特徴とした請求項２０ないし２５いずれか１項に記載のソフトウェア記録媒体。
【請求項２７】
前記複数の手続きに、
前記第２データ内に表現された少なくとも１つの第２集合の少なくとも１つの部分集合をとり前記第３データ内に該部分集合を表現することにより該第３データを誘導することを特徴とする第４手続きをさらに含むことを特徴とした請求項２０ないし２６いずれか１項に記載のソフトウェア記録媒体。
【請求項２８】
前記原始写像が、
恒等写像、定数写像、等号写像、積写像、複数の写像の積写像を与える写像、引戻し演算写像、射影写像、対角写像、置換写像、写像合成写像、評価写像、複数の低位写像を組み合わせて高位写像を与える写像、ＣＵＲＲＹ写像、論理演算写像、ベクトル演算写像、順序写像、汎関数演算写像、固定点演算写像の１つ以上を含むことを特徴とする請求項２０ないし２７いずれか１項に記載のソフトウェア記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明はデータ解析、特にパターンが発見できるようにデータを配置する方法と装置に関する。
【背景技術】
【０００２】
データ管理、データ処理、そしてデータ解析は現代生活及び仕事の上での偏在的要素となっている。科学的、医学的、工学的、そして商業的用途のための膨大なデータの流れの開発、管理、そして保管は、巨大産業となった。バイオテクノロジー、金融、画像、その他のデータのソース、及び需要は、急速に増大している。必ずしもどれが興味ある現象に関係あるか知らないまま、体系的に多数の測定が行われ、大量のデータが自動的に集められている。
【０００３】
したがって、適切な情報を巨大なデータの山から捻り出し、干し草の山の中の針を探すことが益々重要になっている。これは、現在データ解析で使われている多くの技術の背後にある古い仮定と重大な違いを持つ。これらの技術の多くは、例えば科学的知識により前もって丁度適切な変数を測定するなどして精選された、少数の変数を対象にすることを前提

にしている。

【０００４】

これらの技術で使われている基本的な方法論は、常に適用可能では既になくなっている。データ解析における以前の方法の基礎をなす理論はデータ要素の数が個別データの次元より遥かに大きいことを前提としている。しかし、今日ではしばしば、データの次元はデータ要素の数より遥かに大きい。そのような場合はもう例外ではなくある意味で常態である。多くの種類の事象について、その事象を数量化する潜在的に非常に多数の測定可能な項目が存在し、その事象の例の数は比較的少ない。一例を挙げれば、多数の遺伝子と、ある遺伝病に罹患する比較的少数の患者の例がある。他例には画像がある。画像は軽く百万次元（画素）を持ちうるが、一組の解析すべきデータとして百万個の画像を処理することは稀である。

【発明の開示】

【発明が解決しようとする課題】

【０００５】

従って、高次元データがより効果的に解析されうるように与えられたデータを配置する方法と装置を与えるのが本発明の目的である。与えられたデータ内によりうまくパターンを見つけられるようにそのデータを配置する方法を与えるのも本発明の目的である。

【課題を解決するための手段】

【０００６】

本方法は与えられたデータ内にパターンが発見できるようにそのデータを配置することを可能とする。データを特徴づける写像及びそれが所属する集合を利用して、本方法は多数の「データ項目」を比較的少数の入力データ項目から作り、統計的その他の既存のデータ解析手法を適用することを可能にする。データ又はその一部から写像の集合が決定される。次に、既存の写像を組合せるか、ある種の変換を写像に加えることにより、新しい写像が生成される。次に、それらの写像をデータに適用した結果が調べられパターンが探される。例えば、本発明のある実施形態においては、特定の結果データあるいはデータの集合の頻度が調べられる。場合により随意に、ある強いパターンが選ばれ、理想化され、そのパターンを反映したデータを見つけるために伝播され戻される。すなわち、本発明のパターン解析方法は記憶手段と処理手段を備えた情報処理システムにより実行されるパターン解析方法であって、少なくとも１つの第１データを受信する第１工程と、該第１データを前記記憶手段に記録する第２工程と、該第１データの構造と解釈に応じて少なくとも１つの原始写像を決定する第３工程と、該原始写像を前記記憶手段に記録する第４工程と、前記記憶手段に記録された少なくとも１つの第２データを選択する第５工程と、該第２データから少なくとも１つの第３データを誘導する複数の手続きから少なくとも１つの手続きを選択する第６工程と、該手続きに従い前記第２データから前記第３データを誘導する第７工程と、前記記憶手段に該第３データを記録する第８工程とを有し、前記複数の手続きに、前記記憶手段に記録された少なくとも１つの第１写像を所定の方法により選択し該第１写像を前記第２データに適用することにより前記第３データを誘導することを特徴とする第１手続きと、前記第２データ内に表現された複数の集合の直積を少なくとも１つとり該直積を前記第３データ内に表現することにより該第３データを誘導することを特徴とする第２手続きとを含むことを特徴とする。また、前記第９工程は前記第４データのエントロピーに従って前記第１パターンを探すこととすることもできる。また、前記第１パターンが見つかれば該第１パターンを前記記憶手段に記録する第１０工程と、前記記憶手段に記録された少なくとも１つの第２パターンをパターン解析結果として提供する第１１工程をさらに有することもできる。また、前記第１パターンが見つかれば該第１パターンに対応する少なくとも１つの第５データを所定の方法により生成し前記記憶手段に記録する第１２工程をさらに有することもできる。また、前記第１２工程は、前記第４データ内に表現された少なくとも１つの第１確率測度を選択しエントロピーのより低い第２確率測度を該第１確率測度から生成し該第２確率測度を前記第５データ内に表現するか、前記第４データ内に表現された少なくとも１つの第３確率測度を選択し該第３確率測度を集中させて

少なくとも１つの第４確率測度を生成し該第４確率測度を前記第５データ内に表現するか、前記第４データ内に表現された少なくとも１つの第５確率測度を選択し該第５確率測度内の少なくとも１つの確率の集中に各々対応した複数の確率測度を生成し該複数の確率測度を前記第５データ内に表現するか、前記第４データ内の近似的に繰り返すパターンをより正確に前記第５データ内に繰り返させるかの少なくとも１つにより前記第５データを生成することとすることもできる。また、前記第１パターンが見つかれば前記第４データに結び付けられた手続き及び写像に従ってパターン写像を決定し前記記憶手段に記録する第１３工程をさらに有し、前記第８工程は選択された前記手続きと該手続きで写像が使用されていれば該写像を前記第３データと結び付けて前記記憶手段に記録することとすることもできる。また、一連の工程を所定の条件が満たされるまで繰り返す第１４工程をさらに有し、該一連の工程は前記第５ないし１４工程のうち前記第１４工程以外の少なくとも１つを含むこととすることもできる。また、以上において、前記複数の手続きに、前記記憶手段に記録された少なくとも１つの第２写像を選択し前記第２データ内に表現された少なくとも１つの第１集合の該第２写像による逆像をとり前記第３データ内に該逆像を表現することにより該第３データを誘導することを特徴とする第３手続きをさらに含むこともできる。また、前記複数の手続きに、前記第２データ内に表現された少なくとも１つの第２集合の少なくとも１つの部分集合をとり前記第３データ内に該部分集合を表現することにより該第３データを誘導することを特徴とする第４手続きをさらに含むこともできる。また、前記原始写像が、恒等写像、定数写像、等号写像、積写像、複数の写像の積写像を与える写像、引戻し演算写像、射影写像、対角写像、置換写像、写像合成写像、評価写像、複数の低位写像を組み合わせて高位写像を与える写像、ＣＵＲＲＹ写像、論理演算写像、ベクトル演算写像、順序写像、汎関数演算写像、固定点演算写像の１つ以上を含むこととすることもできる。

【０００７】

本発明のいくつかの側面の基本的理解を与えるために、以下に本発明の単純化された要約を示す。この要約は本発明の広範囲にわたる概観ではなく、本発明の鍵となるあるいは決定的な要素を指示することや、本発明の範囲を定めることも意図していない。その唯一の目的は、後のより詳細な記述への前触れとして、本発明のいくつかの概念を単純化した態様で示すことにある。

データ

【０００８】

図１にデータ内にパターンを発見する方法のフローチャートを示す。本方法によれば、まず解析されるべきデータが受信される（１０１）。最も一般的なデータの形態は、遍在する情報処理システムや機器で使われるようなビットの列である。データは通常、何らかの構造と解釈を持つ。例えば、データのある部分は８ビットのグループ毎に一文字と解釈されるテキストデータかもしれない。他の部分は３２ビット整数や６４ビット浮動小数点数を表すかもしれない。あるいは単一のビットが「ｙｅｓ」又は「ｎｏ」と解釈されるかもしれない。遺伝子配列を表すデータでは、２ビットでヌクレオチド中の塩基（Ａ，Ｇ，Ｃ，Ｔのどれか）を表すかもしれない。データはそれぞれが一組の情報を表すいくつかのレコードに分割されている場合もある。例えば画像データは画素数（幅と高さ）を表す２つの整数と、各画素の色を表す整数の列からなるかもしれない。

【０００９】

表記

【００１０】

以下では、データをもう少し抽象的に取り扱う。整数は、それを表すのに何ビット使われていようが整数と呼ぶ。同様に、浮動小数点数のことは実数と呼び、「ｙｅｓ」と「ｎｏ」のように二者択一を表すデータは全てブール値と呼ぶ。さらに一般に、以下では各種の集合と写像について言及する。

【００１１】

集合は要素の集まりである。例えば、整数の集合Ｚは全ての整数を要素に持つ集合である

。ブール値の集合ｂｏｏｌはｔｒｕｅとｆａｌｓｅの２つの要素しか持たない。集合はその全ての要素を「｛｝」内に列挙して表記されることもある。例えばｂｏｏｌ＝｛ｔｒｕｅ，ｆａｌｓｅ｝のように。表記ａ∈Ａはａが集合Ａの要素であることを表す。もし集合Ｂの全ての要素がもう１つの集合Ａの要素でもあるならば、ＢはＡの部分集合であり、これをＡ⊃Ｂ（又はＢ⊂Ａ）と表記する。二つの集合ＡとＢは、もしＡ⊃ＢかつＢ⊃Ａならば等しい（Ａ＝Ｂと表記）。Ａの部分集合Ｂは、もしＡ≠Ｂならば真部分集合である。

【００１２】

これらの表記を使うことは、本発明が実際に集合という数学的概念を扱うことを意味しない。それは本方法を簡潔に、（これらの表記が概念の記述に、しばしばあまり厳密でなく、使われる）関連技術分野で熟練した者によく知られた表記で記述するためである。例えば、Ｚのように無限個の要素を持つ集合があり、また（実数のように）正確に指定するには無限の精度を要する要素を持つ集合もあるが、それらは有限の存在である情報システムで日常的に扱われている。これは、普通はそのような集合の有限個の要素しか、各仕事に必要ないからである。また集合は時に記号的に処理され、または近似されることもある。集合や写像を表現し操作するこれらや他の手法は関連技術分野である計算機科学では良く知られている。ＳＥＴＬやＭＩＲＡＮＤＡ等のある種のプログラム言語は集合を言語プリミティブとして持つほどである。また、ここで使われる集合や写像の概念はＭＬやＨＡＳＫＥＬのような型付き関数型言語における型と写像の概念に非常に近い。従って、関連技術分野における通常の能力を持つ者は適切な手法を使ってここに開示される本方法を実現することができるであろう。

【００１３】

集合ＡとＢに対して、「Ａ→Ｂ」はＡからＢへの写像の集合を表す。写像とは与えられた集合の各要素に一意の対象を結びつける方法である。つまりＡからＢへの写像とは、Ａの各要素ａにただ一つのＢ内の対象ｆ（ａ）を与えるような関数のことである。そのような状態は時に「ｆはａをｆ（ａ）に送る（あるいは写像する）」と記述される。表記「ｆ：Ａ→Ｂ」はｆが集合Ａから集合Ｂへの写像であること、即ちｆがＡ→Ｂの要素であることを意味する。写像ｆ：Ａ→Ｂに対して、Ａはｆの定義域と呼ばれる。

【００１４】

集合Ａについて、id_A：Ａ→ＡはＡの各要素ａをそれ白身に送る恒等写像を表す。

【００１５】

集合ＡとＢについて、定数写像ｃｏｎｓｔ：Ａ→（Ｂ→Ａ）はｃｏｎｓｔ（ａ）（ｂ）＝ａで定義される。つまり、Ａの要素ａについて、ｃｏｎｓｔ（ａ）：Ｂ→ＡはＢのどんな要素ｂもａに送る写像である。

【００１６】

ＢがＡの部分集合である時、包含写像ｉｎｃｌ：Ｂ→Ａはｉｎｃｌ（ｂ）＝ｂで定義される。

【００１７】

２つの集合ＡとＢについて、Ａ×Ｂはこの２集合の直積、つまり順序対（ａ，ｂ）（ａはＡに、ｂはＢに属する）の集合を表す。同様に、Ａ×Ｂ×Ｃは３集合Ａ、Ｂ、Ｃの直積を表し、以下同様である。一般に別の集合Ｉでインデックスされた任意の集合族A_iの直積は、$\Pi_{i \in I} A_i$または、A_iが全て等しい時にはA^Iで表される。$\Pi_{i \in I} A_i$の要素は$(a_i)_{i \in I}$で表される。ここで各a_iはA_iの要素である。有限個の要素を持つ標準集合を次のように表記する。$Z_1 = \{1\}$、$Z_2 = \{1, 2\}$、．．．、$Z_n = \{1, \ldots, n\}$。以下では、Ａ×Ｂは$I = Z_2$、$A_1 = A$、$A_2 = B$のときの$\Pi_{i \in I} A_i$の略記と理解されたい。同様に、Ａ×Ｂ×Ｃは$I = Z_3$、$A_1 = A$、$A_2 = B$、$A_3 = C$のときの$\Pi_{i \in I} A_i$の略記であり、以下同様である。

【００１８】

写像ｆ：Ａ→Ｂは、各ａ∈Ａについてｆのａ番目の要素をｆ（ａ）と考えることによって、B^A即ちＡでインデックスづけされたＢのコピーの直積の要素と考えられる。したがって、Ａ→ＢはここではB^Aの別名とみなされる。

【００１９】
特別な集合ｕｎｉｔが定義される。それはただ１つの要素を持つ。ｕｎｉｔによって、集合Ａの任意の要素ａを、ｕｎｉｔの唯一の要素をａに送る写像ａ：ｕｎｉｔ→Ａとみなすことができる。写像に対してのみに適用可能な写像または操作を集合Ａの（写像でない）普通の要素に適用するために、本発明が自動的にこの変換を実行することもある。A^{unit}あるいはｕｎｉｔ→Ａという形の集合はＡと同一視される。
【００２０】
写像ｆ：Ａ→ＢとＢの要素ｂについて、ｆによるｂの逆像f^{-1}（ｂ）とは、ｆによりｂに送られるＡの要素からなるＡの部分集合である。Ｂの部分集合Ｃのｆによる逆像f^{-1}（Ｃ）とは、ｆによりＣの要素に送られるＡの要素からなるＡの部分集合である。
【００２１】
ある種の写像は再帰的に定義される。つまり、再帰的に定義された写像はその定義にそれ自身を使用する。例えば、階乗関数ｆａｃ：Ｎ→Ｎは自然数ｎを、もしｎが１ならば１に、それ以外ならばｆａｃ（ｎ）のｎ倍に送る関数として定義される（ここでＮは自然数の集合｛１，２，３，．．．｝を表す）。
【００２２】
引戻し
【００２３】
２つの積集合$\Pi_{i \in I} A_i$と$\Pi_{j \in J} B_i$について、全てのｊ∈ＪについてＡ$_{h(j)}$＝Ｂ$_j$である写像ｈ：Ｊ→Ｉが存在する時、対応する引戻し$h^*：\Pi_{i \in I} A_i \to \Pi_{j \in J} B_i$が$(h^*((a_i)_{i \in I}))_j = a_{h(j)}$で定義される。この写像の特別な場合に以下がある。
【００２４】
［ＰＢ１］Ｉの任意の部分集合Ｊについて、ｈ＝ｉｎｃｌ：Ｊ→Ｉとした$h^*：\Pi_{i \in I} A_i \to \Pi_{j \in J} A_j$は射影写像を定義する。例えば直積Ａ×Ｂについて、自然な射影がある：
・ｐｒｏｊ$_A$：Ａ×Ｂ→Ａ　［ｐｒｏｊ$_A$（ａ，ｂ）＝ａ］
・ｐｒｏｊ$_B$：Ａ×Ｂ→Ｂ　［ｐｒｏｊ$_B$（ａ，ｂ）＝ｂ］
写像ｐｒｏｊ$_A$は$h^*：\Pi_{i \in Z_2} A_i \to \Pi_{j \in Z_1} B_j$でＡ$_1$＝Ａ、Ａ$_2$＝Ｂ、ｈ＝ｉｎｃｌ：Ｚ$_1$→Ｚ$_2$としたものと同じである。
【００２５】
［ＰＢ２］同じ集合ｎ個のコピーの直積Ａ×Ａ×．．．×Ａについて、対角写像ｄｉａｇ：Ａ→Ａ×Ａ×．．．×Ａがｄｉａｇ（ａ）＝（ａ，ａ，．．．，ａ）で定義される。これは$h^*：\Pi_{i \in Z_1} A_i \to \Pi_{j \in Z_n} B_j$でＡ$_1$＝Ａ、Ｂ$_j$＝Ａとし、ｈ：Ｚ$_n$→Ｚ$_1$をＺ$_n$＝｛１，．．．，ｎ｝の全てのｊについてｈ（ｊ）＝１で定義したものと同じである。
【００２６】
［ＰＢ３］直積Ａ×Ｂについて、（ａ，ｂ）を（ｂ，ａ）に送る交換写像Ａ×Ｂ→Ｂ×Ａがある。同様に任意の数の集合の直積について、成分の順序を変える置換写像がある。これは$h^*：\Pi_{i \in Z_n} A_i \to \Pi_{j \in Z_n} B_j$でｈを置換写像としＺ$_n$＝｛１，．．．，ｎ｝の全てのｊについてＢ$_j$＝Ａとしたものと同じである。
【００２７】
［ＰＢ４］２つの写像ｆ：Ａ→Ｂとｇ：Ｂ→Ｃについて、合成写像ｇ^{o}ｆ：Ａ→ＣがＡ内のａについてｇ^{o}ｆ（ａ）＝ｇ（ｆ（ａ））で定義される。これも引戻しの特別な場合である。これをみるには全てのＣ$_b$とＣ$_a$をＣと等しくしてｇ∈$C^B = \Pi_{b \in B} C_b$でありｇ^{o}ｆ∈$C^A = \Pi_{a \in A} C_a$であることを思い出されたい。
【００２８】
［ＰＢ５］集合ＡとＢ、Ａ内のａについてｃｏｎｓｔ（ａ）：Ｂ→ＡはＢ内の任意のｂをａに送る写像である。Ｊ＝Ｚ$_n$として定数写像ｃｏｎｓｔ（ａ）：Ｊ→Ａとその引戻しｃｏｎｓｔ（ａ）$^*：\Pi_{i \in A} B \to \Pi_{j \in J} B$を考える。それは写像ｆ：Ａ→Ｂを、そのａ

での値ｆ（ａ）∈Ｂに写像する。これは、ｅｖ（ｆ，ａ）＝ｆ（ａ）で定義される、写像の値を評価する写像ｅｖ：（Ａ→Ｂ）×Ａ→Ｂを定義する。

【００２９】

統計

【００３０】

本発明においては、データを確率測度（確率分布）のような統計として表現すること、あるいはもっと一般に、データの相対頻度を処理することが、特に有用である。一般に、集合Ａについて、Ａ上の確率測度ＰｒはＡの（事象と呼ばれる）部分集合Ｂに対して０と１の間の実数Ｐｒ（Ｂ）を与える。データを確率測度で表すとは以下を意味する。もしあるデータが集合Ａの単一の要素ａであるなら、それはＡの事象ＢがａをＰｒ（Ｂ）＝１を与え、それ以外のときＰｒ（Ｂ）＝０を与える確率測度として表現されうる。あるいはそれはａを中心としたガウス分布のような、概算測度としても表現されうる。同じ集合に属する多くのデータ点があるときには、Ａに含まれる全てのデータ点に対するＢに含まれるデータ点の比を与える、単純な係数測度Ｐｒ（Ｂ）として表現されるかもしれないし、あるいは再び、ガウス混合分布やＰａｒｚｅｎ　Ｗｉｎｄｏｗの手法のような概算測度としても表現されうる。情報システムにおけるそのような確率測度の処理及びシミュレーションのための種々の手法が、関連技術分野ではよく知られている。後述のある実施形態においては、頻度係数と呼ばれる具体的な方法が使われる。このように確率測度を使うとき、各週強情の標準測度が必要に応じて使われる。これは、一様分布のように、その集合の、特徴のないデフォルト状態を表す確率測度である。

原始写像

【００３１】

次に、そのデータかそのデータの一部からの写像の集合が決定される（１０２）。これらの写像は原始写像と呼ばれる。原始写像に含まれる写像は集合上に定義される標準写像の一つかもしれない。例えば、整数の集合Ｚには、ある整数をその次に数に送る、自身への写像がある。集合Ｚにはまた加法もある。それはＺ×ＺからＺへの写像として表現さるのだが、これも原始写像の集合に加えられるかもしれない。このように加法写像はＺ×Ｚ内の（ｉ，ｊ）をＺ内のｉ＋ｊに送る。従って、データの一部が一つあるいは複数の整数を表していれば、その整数の次の数を与える写像あるいはそれらの整数の和を与える写像が原始写像に含められるかもしれない。ある種の集合らはそれらの間に自然な写像を持つ。例えば、任意の集合Ａについて、等しさという概念はＡ×Ａからブール値の集合ｂｏｏｌ＝｛ｔｒｕｅ，ｆａｌｓｅ｝への写像を定義する。つまり、Ａ×Ａ内の（ｕ，ν）に対してその写像はｕ＝νであるときに限りｔｒｕｅを与える。同様に、ある種の集合には順序の概念があり、写像と考えることができる。例えば整数の集合Ｚに、Ｚ×Ｚ内の（ｉ，ｊ）にｉ＜ｊのときに限りｔｒｕｅを与えるＺ×Ｚからｂｏｏｌへの順序写像がある。

【００３２】

以下に集合に自然に随伴し、原始写像の集合に含められるかもしれない写像のいくつかを列挙する。ここでＲは実数の集合を表す。

【００３３】

［ＰＭ　Ｉ］任意の集合Ａは次の原始写像を持つ：

・恒等写像：id_A：Ａ→Ａ　［id_A（ａ）＝ａ］

・定数写像：ｃｏｎｓｔ：Ａ→（Ｂ→Ａ）［ｃｏｎｓｔ（ａ）（ｂ）＝ａ］（任意の集合Ｂについて）

【００３４】

［ＰＭ　ＩＩ］等しいかどうか簡単に決定できる集合Ａについて、等号写像：

・eq_A：Ａ×Ａ→ｂｏｏｌ　［ａ＝ｂならeq_A（ａ，ｂ）＝ｔｒｕｅ、それ以外ｆａｌｓｅ］

【００３５】

［ＰＭ　ＩＩＩ］２つの写像ｆ：Ａ→Ｂとｇ：Ｃ→Ｄについて、積写像ｆ×ｇ：Ａ×Ｃ→Ｂ×Ｄがｆ×ｇ（（ａ，ｃ））＝（ｆ（ａ），ｇ（ｃ））で定義される。これは原始写像

ｍｐ：（Ａ→Ｂ）×（Ｃ→Ｄ）→（Ａ×Ｃ→Ｂ×Ｄ）を定義する。

【００３６】

［ＰＭ　ＩＶ］写像に対する引戻し：ｐｕｌｌｂａｃｋ：（Ｊ→Ｄ→（$\Pi_{i \in I} A_i \rightarrow \Pi_{j \in J} B_i$）。これは写像を別の写像に送る。これの特別な場合には射影写像［ＰＢ１］、対角写像［ＰＢ２］、置換写像［ＰＢ３］、写像合成写像［ＰＢ４］、評価写像［ＰＢ５］が含まれる。

【００３７】

［ＰＭ　Ｖ］低位写像の組み合わせ。Ｋをインデックスの集合とし、各ｋ∈ＫについてＩ$_k$もインデックスの集合とする。ｋ∈Ｋについて既知の写像$f_k : \Pi_{i \in I_k} A_{k,i} \rightarrow B_k$があり、もう一つのインデックス集合Ｊと、$A_{k,i} \neq A_{m,j}$のとき$h_k(i) \neq h_m(j)$であるような写像ｈ：Ｉ→Ｊもあると仮定する。写像$F : \Pi_{k \in K} \Pi_{i \in I_k} A_{k,i} \rightarrow \Pi_{k \in K} B_k$とｈ：Ｌ→Ｊを、ＦはＫ内の全てのｋについての$f_k$の積集合として、$L = U_{k \in K} I_k$はインデックス集合$I_k$の共通部分のない和集合として、そしてｈはＩ$_k$上で$h_k$と一致するように、それぞれ定義する。すると、ｈの引戻し$h^* : \Pi_{j \in J} A_j \rightarrow \Pi_{k \in K} \Pi_{i \in I_k} A_{k,i}$とＦを合成すると新しい写像$F \circ h^* : \Pi_{j \in J} A_j \rightarrow \Pi_{k \in K} B_k$が定義される。

【００３８】

［ＰＭ　ＶＩ］ｃｕｒｒｙ写像ｃｕｒｒｙ：（Ａ×Ｂ→Ｃ）→（Ａ→（Ｂ→Ｃ））は、写像ｆ：Ａ×Ｂ→Ｃを、写像ｃｕｒｒｙ（ｆ）：Ａ→（Ｂ→Ｃ）に送るが、これはＡ内のａをｃｕｒｒｙ（ｆ）（ａ）（ｂ）＝ｆ（ａ，ｂ）で定義される写像ｃｕｒｒｙ（ｆ）（ａ）：Ｂ→Ｃに送る。逆の操作はｕｎｃｕｒｒｙ写像ｕｎｃｕｒｒｙ：（Ａ→（Ｂ→Ｃ））→（Ａ×Ｂ→Ｃ）で、これは写像ｇ：Ａ→（Ｂ→Ｃ）を、（ａ，ｂ）∈Ａ×Ｂをｇ（ａ）（ｂ）に送る別の写像ｕｎｃｕｒｒｙ（ｇ）：Ａ×Ｂ→Ｃに送る。これは計算機科学ではよく知られている。

【００３９】

［ＰＭ　ＶＩＩ］各種の論理演算がある：ＮＯＴ：ｂｏｏｌ→ｂｏｏｌ、ＡＮＤ：ｂｏｏｌ×ｂｏｏｌ→ｂｏｏｌ、ＯＲ：ｂｏｏｌ×ｂｏｏｌ→ｂｏｏｌ等。

【００４０】

［ＰＭ　ＶＩＩＩ］Ｒを含む任意のベクトル空間は次の自然な写像を持つ：

・（加法）$Add_V : V \times V \rightarrow V$　［$Add_V(u, v) = u + v$］

・（実数との積）$Mult_V : R \times V \rightarrow V$　［$Mult_V\{a, v) = av$］

・（減法）$Sub_V : V \times V \rightarrow V$　［$Sub_V(u, v) = u - v$］　（これは加法と－１倍によって定義できるが、後の記法の簡略化のためにここに含める。）

・（長さ）$Len_V : V \rightarrow R$　［$Len_V(v)$＝ベクトルｖの長さ］

・別のベクトル空間でパラメーターづけられた種々の線形変換：ＬＴ：Ｖ×Ｕ→Ｗ

・別のベクトル空間でパラメーターづけられた種々の双線形、ｔｒｉｌｉｎｅａｒ、．．．等々の形式：

　・ＬＦ：Ｖ×Ｕ→Ｒ

　・ＢＦ：Ｖ×Ｖ×Ｕ→Ｒ

　・ＴＦ：Ｖ×Ｖ×Ｖ×Ｕ→Ｒ

【００４１】

［ＰＭ　ＩＸ］Ｒは順序の概念を持つ：

・$Ord_R : R \times R \rightarrow bool$　［ａ＜ｂならＯｒｄ$_R$（ａ，ｂ）＝ｔｒｕｅ、それ以外はｆａｌｓｅ］

【００４２】

［ＰＭ　Ｘ］ユークリッド空間Ｅは２点間のベクトルの概念を持つ：

・$Diff_E : E \times E \rightarrow V$（Ｖは同次元のベクトル空間）

【００４３】

［ＰＭ　ＸＩ］Ｒの部分集合Ａ上の実数値関数のある種の集合Ｕ（つまりＵはＡ→Ｒの部分集合）について、微分写像Ｄｅｒ：Ｕ→（Ａ→Ｒ）は関数をその導関数（微分）に送る

。実ベクトル空間の間の写像の様々な微分をとる同様な写像がある。さらに一般に、原始写像として加えられるかもしれないよく知られた数学的変換は他にもある（例えばフーリエ変換）。

【００４４】

［ＰＭ　ＸＩＩ］固定点演算。写像ｆ：Ａ→Ａについて、固定点演算子Ｆｉｘ：（Ａ→Ａ）→Ａはその写像のある固定点を与える。つまり、ａ＝Ｆｉｘ（ｆ）はｆ（ａ）＝ａであるようなＡの要素である。これは、再帰的に定義される写像を定義するのに使える。例えば、上述の階乗写像ｆａｃ：Ｎ→Ｎを再帰的でない写像から得ることができる。写像ｆ：Ｎ→Ｎを別の写像Ｆ（ｆ）：Ｎ→Ｎに送る写像Ｆ：（Ｎ→Ｎ）→（Ｎ→Ｎ）を次のように定義する。Ｆ（ｆ）は自然数ｎをｎ＝１なら１に、それ以外ならｆ（ｎ－１）のｎ倍に送る。このとき、Ｆｉｘ（Ｆ）が階乗写像である。固定点演算は全ての写像に適用可能ではないかもしれないことに注意せよ。

【００４５】

原始写像はまた、表現されたデータにもっと特有のものであるかもしれない。もしデータ中のある整数がある人の課税所得を表すなら、その所得に対する税額を与える写像も、アプリケーションの必要に応じて、原始写像として含められるかもしれない。

誘導データと写像

【００４６】

次のステップ（１０３）では、そのデータと原始写像をもとに、他のデータや写像が生成される。これらの生成法のうちのいくつかは以下の通り。

・２つ以上の集合から積集合を作られ得る。積集合上の確率測度は元の集合上のものから誘導され得る。

・データは写像によって送られ得る。確率測度は写像によって誘導され得る。

・集合の写像による逆像がとられ得る。

・データは部分集合に制限され得る。確率測度も部分集合に制限され得る。

・写像を別の写像に送る写像が適用されて、新しい写像が作られ得る、例えば：

・２つの写像ｆ：Ａ→Ｂとｇ：Ｃ→Ｄから、積写像ｆ×ｇ：Ａ×Ｃ→Ｂ×Ｄがｆ×ｇ（（ａ，ｃ））＝（ｆ（ａ），ｇ（ｃ））で定義される。（［ＰＭ　ＩＩＩ］参照）

・２つの写像ｆ：Ａ→Ｂとｇ：Ｂ→Ｃから、写像ｇ°ｆ：Ａ→ＣがＡ内のａについて（ｇ°ｆ）（ａ）＝ｇ（ｆ（ａ））で定義される。（［ＰＭ　ＩＶ］参照）

・より高位の写像、つまり引数のより多い写像は、多くの対象の間の関係を定義するため重要である。写像を組み合わせてより高位の写像に導くことは、原始写像の殆どは多くても２つの引数しか持たないことから、特に重要である。このように、［ＰＭ　Ｖ］の原始写像は重要である。それは上述した写像の写像を適用する特殊な場合に過ぎないが、ここで例を使って簡単に説明する価値がある。ｆ：Ａ×Ａ→Ｂを写像とする。高位の写像を作るために、まず積写像を作る：ｆ×ｆ：Ａ×Ａ×Ａ×Ａ→Ｂ×Ｂ。しかしこれは同じことを２回やっているだけだから、あまり多くの新情報をもたらさない。しかし、ｇ（ａ，ｂ，ｃ）＝ｆ×ｆ（ａ，ｂ，ｂ，ｃ）で定義されるｇ：Ａ×Ａ×Ａ→Ｂ×Ｂは３つの引数の間に新しい関係を定義する。これが、［ＰＭ　Ｖ］の原始写像が適用されたときにこの場合に起こることである。

【００４７】

上に列挙したように、本方法の様々な段階で新しいデータと写像を生成するための、方法と源の選び方はたくさんある。アプリケーションと、既に見つかったデータと写像を基に有用なパターンを見つける可能性がよりよくなるように、生成されるデータと写像を選ぶための計画があるべきである。一般に、パターン写像（下記参照）とされた写像は、新しい写像の構成要素として使われるより強い傾向を持つべきである。また、なにかのパターンが見つかった集合は源の集合としてより頻繁に使われるべきである。本発明の実施例で使われている一つの方法を後述する。

パターン

【００４８】

次のステップ（１０４）では、生成された様々なデータと写像の中にパターンが存在するかどうかが調べられる。これは、繰り返されたデータを見つけたり、確率測度の低いエントロピーのような統計的に有意味な条件を追求したり、比較的少数の要素への確率の集中を検出するなどの、パターン発見のための従来手法のいずれでもを使ってなされる。以下では、その中にパターンが見つかったデータをパターンデータと呼ぶ。

【００４９】

パターンデータは元のデータと生成されたデータに何かの写像を適用した結果である。これらの写像を以下ではパターン写像と呼ぶ。パターン写像はパターン解析に重要である。例えばもし写像をデータに適用した結果がおおまかに繰り返すパターンであるとか、あるいはある確率測度からある写像で誘導された確率測度が低いエントロピーを持っているなら、これらの写像は元のデータを何らかの面で特徴付けている。このパターン写像は類似のデータ中に同じ特徴があるかどうか調べるために適用するのに有用であろう。様々なパターン写像の組合せは、元の集合や各中間段階のデータを特徴付けるかもしれない。

【００５０】

パターンの存在を決定するとき、写像そのものから来るものを考慮に入れねばならない。つまり、もし写像そのものがパターンを作るなら、そのパターンはデータの特徴を表さない。例えば、上述のエントロピーは、何のパターンも持たない何か（例えばパターン写像の定義集合上の標準確率分布等）に同じパターン写像を適用した結果と相対的に評価しなければならない。

バックトラック

【００５１】

場合により随意に、次のステップ（１０５）では、本方法は前ステップで見つかったパターンデータをとってそのパターンに対応する「理想的」データを生成し得る。まず、（パターンデータが見つかったのと）同じ集合内に、パターンデータを修正することで新しいデータが作られるかもしれない。もしそのパターンデータが、生成された集合上のエントロピーの低い確率測度として見つけられたならば、さらに低いエントロピーを持った理想化された確率測度がその集合上に導入されるかもしれない。そして、パターン写像を通してその理想化された確率測度を誘導する確率測度が見つけられるかもしれない。もし確率の集中が観察されたのなら、理想化はそれをもっと集中するかもしれない。また、もし比較的少数の集中しかないのなら、それぞれ１つの集中を持つ複数の確率測度が、新しいパターンデータとして作られるかもしれない。おおまかに繰り返すパターンは正確に繰り返すパターンにされるかもしれない。

【００５２】

それから、理想化されたパターンの、対応するパターン写像による逆像がとられるかも知れない。元のデータの入っていた集合まで遡る上での中間段階の集合の中の可能なデータの集合がこうして同定される。これは、そのデータがパターン写像によって理想化されたパターン内に送られたときにｔｒｕｅを与えるその集合上の述語論理を作ることによって実装され得る。また、元のデータのこの集合内にある部分（つまり、対応する述語論理にｔｒｕｅを与えられる部分）は特に重要である。なぜならこの部分的データは他の写像によって前へ送られて他に何かパターンが現われるかどうか調べられ得るからである。

【００５３】

このようにしてパターンを持つ可能なデータの集合が同定できる。十分多くのパターン使い、そのような逆像の共通部分をとることで、可能なデータの小さな集合あるいはただ１つのデータさえ見つかるかもしれない。

【００５４】

次のステップ（１０６）では、望ましいデータが出力される。これは見つかったパターンや、それらに対応する「純粋な」データを含むかもしれない。

【００５５】

最後に、プロセスの停止条件が調べられ（１０７）、もし条件に合わなければプロセスは繰り返す。

【図面の簡単な説明】
【００５６】
【図１】データ内にパターンを見つける方法のフローチャートを示す。
【図２】探索アルゴリズムのフローチャートを示す。
【図３】データ構造ＦＣと、ＦＣ内で使われる部分構造を図式的に表す。
【図４】理想化プロセスのフローチャートを示す。
【発明を実施するための最良の形態】
【００５７】
以下の記述では、本発明の完全な理解を与えるために、説明の目的で多数の特定細部が提示される。しかし、関連技術分野で熟練した者には、本発明がそれらの特定細部なしでも実施可能であることが明確であり得る。他の場合には、本発明の記述のために、よく知られた構造や装置がブロックダイアグラム中に示される。本発明は様々な形態のハードウェア、ソフトウェア、ファームウェア、特殊用途プロセッサ、あるいはそれらの組合せによって実装され得ることが理解されるべきである。好ましくは、本発明はプログラム記憶装置に有体的に有形化されたアプリケーションプログラムのソフトウェアとして実装されるべきである。そのアプリケーションプログラムは、任意の適当なアーキテクチャからなる機械に読み込まれ、実行され得る。好ましくは、その機械は、１つあるいは複数の中央処理装置（ＣＰＵ）、ランダムアクセスメモリ（ＲＡＭ），入出力（Ｉ／Ｏ）インタフェースのようなハードウェアを持つコンピュータプラットフォーム上に実装されるべきである。そのコンピュータプラットフォームはまたオペレーティングシステムとマイクロ命令コードを含む。本明細書に記述された様々なプロセスや関数は、オペレーティングシステムにを通して実行されるそのマイクロ命令コードか、あるいはアプリケーションプログラム、あるいはそれらの組合せであるかもしれない。加えて、追加のデータ記憶装置や印刷装置など、他の様々な周辺装置がそのコンピュータプラットフォームに接続されるかもしれない。さらに理解されるべきことは、付随する図に描かれたシステム構成要素と方法ステップの一部は好ましくはソフトウェアに実装されるべきなので、本発明がプログラムされる態様に依存して、システム構成要素（あるいは方法ステップ）の間の実際の接続は異なるかもしれないことである。本明細書に記述された本発明の教示によれば、関連技術分野の技術者は、これらのあるいは類似した本発明の実装あるいは配置を企図することができるであろう。
データ
【００５８】
ここでは、データを分析するための本発明の実施例を提示する。明快さのために、関連技術分野の技術者にはよく知られる一定の抽象性が維持される。例えば、集合や写像は、情報システム上のデータとして表現、又は情報システム上のデータによって近似される。
【００５９】
頻度あるいは確率が本発明で以下に操作されるかを描写するため、頻度計数というデータ構造がここに開示される。それは集合上の単純な計数確率測度をモデル化するための具体的方法である。本実施例では、全てのデータはある集合上の頻度計数として表される。
【００６０】
以下では、任意の集合Ａについて、Ａ上の頻度計数とは、Ａの要素とその数を捉えるデータを意味する。それは、Ａのいかなる要素も２度以上現れないようなＡ×Ｎの部分集合として扱われる。ここでＮ＝｛１，２，３，．．．｝つまり自然数の集合である。Ａ上の頻度計数の集合はＦｒｅｑ（Ａ）で表される。従ってＡ上の頻度計数即ちＦｒｅｑ（Ａ）の要素Ｆは、Ａの要素ａと自然数ｎの組（ａ，ｎ）の集合Ｆであって、もし（ａ，ｎ）を含めば（ａ，ｍ）の形の他のどんな要素も含まないようなものである。頻度計数内のこれらの組は以下では粒子と呼ばれる。Ａの要素ａとＡ上の頻度計数Ｆについて、$count_F$（ａ）と記述されるａの計数とは、もしＦ内に（ａ，ｎ）の形の要素があればｎで、なければ０で定義される。ｍａｓｓ（Ｆ）すなわちＦのマスは、Ａ内の全てのａについての$count_F$（ａ）の和として定義される。そしてａの確率P_F（ａ）は、$count_F$（

ａ）をｍａｓｓ（Ｆ）で割ったものと定義される。Ｆの台ｓｕｐｐ（Ｆ）は、ｃｏｕｎｔ$_{Ｆ}$（ａ）＞０であるａからなるＡの部分集合と定義される。Ｆのエントロピー Ｈ（Ｆ）は、ｓｕｐｐ（Ｆ）内の全てのａについての和、$-\Sigma_{a\in supp(F)}P_F(a)\log_2 P_F(a)$で定義される。

【００６１】

後の参考のために次に注意すべきである。

【００６２】

［ＦＣ　Ｉ］２つの頻度計数、Ａ上のＦとＢ上のＧから、Ａ×Ｂ上の（直積）頻度計数Ｆ×Ｇが次のように生成できる。Ｆ×Ｇは、Ｆ内の粒子（ａ，ｎ）とＧ内の粒子（ｂ，ｍ）の全ての組合せについて（（ａ，ｂ），ｎｍ）という粒子を持つ（Ａ×Ｂ）×Ｎの部分集合である。これは直積確率測度に対応する。

【００６３】

［ＦＣ　ＩＩ］写像ｆ：Ａ→Ｂがあるとき、頻度計数の写像f_*：Ｆｒｅｑ（Ａ）→Ｆｒｅｑ（Ｂ）が次のように定義される。頻度計数Ｆに対して、f_*（Ｆ）は、ｂ＝ｆ（ａ）なる粒子（ａ，ｍ）が少なくとも１つＦ内に存在しｎはそのような粒子（ａ，ｍ）全てについてのｍの和であるような粒子（ｂ，ｎ）からなる。言い換えれば、集合f_*（Ｆ）は、Ｆ内の全ての（ａ，ｍ）について（ｆ（ａ），ｍ）を追加し、その後に、同じ第一成分を持つ異なる粒子がなくなるまで、同じｂの（ｂ，ｉ）と（ｂ，ｊ）を（ｂ，ｉ＋ｊ）で置き換えてゆくことで作られる。これは誘導された確率測度に対応する。

【００６４】

［ＦＣ　ＩＩＩ］もしＡ⊃ＢならばＦｒｅｑ（Ａ）⊃Ｆｒｅｑ（Ｂ）である。つまり、Ｂ上の頻度計数は自動的にＡ上の頻度計数である。Ａ⊃ＢでＦがＡ上の頻度計数であるとき、ＦのＢへの制限$F|_B$とは、Ｆの粒子（ａ，ｎ）でａがＢに含まれるもの全てからなるＢ上の（従ってＡ上の）頻度計数である。

【００６５】

［ＦＣ　ＩＶ］Ａ上の２つの頻度計数ＦとＧは、ある数ｍ＞０があって、Ａの全てのａについて$count_F(a)=m\ count_G(a)$であるとき、同値であるといわれる。もしＦとＧが同値なら、様々な性質がある。ｍａｓｓ（Ｆ）＝ｍ　ｍａｓｓ（Ｇ）、ｓｕｐｐ（Ｆ）＝ｓｕｐｐ（Ｇ）、Ａの全てのａについて$P_F(a)=P_G(a)$、そしてＨ（Ｆ）＝Ｈ（Ｇ）。

【００６６】

［ＦＣ　Ｖ］集合Ａについて、Ａ上の標準頻度計数Ｓｔ（Ａ）は、Ａ内の各ａについて粒子（ａ，ｌ）を持つＡ×Ｎの部分集合として定義される。この定義と［ＦＣ　Ｉ］によれば、Ｓｔ（Ａ）×Ｓｔ（Ｂ）はＳｔ（Ａ×Ｂ）と同一であることに注意せよ。

原始写像

【００６７】

［ＰＭ　Ｉ］以下に列挙された全ての原始写像が、原始写像の集合に含まれる。

誘導データと写像

【００６８】

ロードされたデータと原始写像に基いて、そのデータを特徴付ける様々な集合の可能性を探索するために、他のデータと写像が生成される。始めは、入力データが集合上の頻度計数として表現されたものがある。従ってシステムはその集合に適用可能な写像を試しに適用することから始める。そのような写像を適用した結果は新しいデータである。具体的には、プロセスは次のようなデータ構造を維持する：

・頻度計数の表現を格納するデータ構造ＦＣ。それは始め、頻度計数として表現された入力データと、入力データがその上にあるような集合の構成要素として現れる全ての集合Ａについて、その標準頻度計数Ｓｔ（Ａ）（［ＦＣ　Ｖ］参照）を持つ。（つまり、もし入力データがＡ×（Ｂ→Ｃ）上の頻度計数なら、Ａ，Ｂ，Ｃ，Ｂ→Ｃ、Ａ×（Ｂ→Ｃ）上の標準頻度計数がＦＣ内に含まれるであろう。）それはまたｂｏｏｌやｕｎｉｔなどのいくつかの標準的集合上の標準頻度計数も含む。

・集合の記号表現を格納するデータ構造ＳＥＴＳ。始めそれはＦＣ内の頻度計数がその上にあるような集合を含む。
・写像の記号表現を格納するデータ構造ＭＡＰＳ。始めそれは原始写像を含む。
【００６９】
過程が続く上で、ＦＣ、ＳＥＴＳ、ＭＡＰＳに、以下の何れかの様に要素が加えられる。
【００７０】
［Ｄ　Ｉ］もしＦＣ内に頻度計数の組Ｆ、Ｇがあれば、Ｆ×ＧがＦＣに加えられ得る（［ＦＣ　Ｉ］参照）。３つ以上の頻度計数の組についても同様である。
【００７１】
［Ｄ　ＩＩ］もしＭＡＰＳ内の写像がＭＡＰＳ内の写像に適用できるなら（例えば［ＰＭ　ＩＩＩ］、［ＰＭ　ＩＶ］、［ＰＭ　Ｖ］、［ＰＭ　ＶＩ］、［ＰＭ　ＸＩＩ］）、適用した結果の写像がＭＡＰＳに加えられ得る。例えば、いくつかの写像の組が選ばれてそれらの積写像あるいは可能ならそれらの合成がＭＡＰＳに加えられ得る。あるいは任意の写像が他の写像に適用されて結果がＭＡＰＳに加えられ得る。
【００７２】
［Ｄ　ＩＩＩ］ＳＥＴＳ内の集合の部分集合がＳＥＴＳに加えられ得る。頻度計数が部分集合に制限されうる。部分集合の逆像がＳＥＴＳに加えられ得る。Ａの部分集合Ｂについて、部分集合判別写像$subset_B$：Ａ→ｂｏｏｌ（ａ∈Ｂなら$subset_B$（ａ）＝ｔｒｕｅ、それ以外ならｆａｌｓｅと定義される）がＭＡＰＳに加えられ得る。
【００７３】
［Ｄ　ＩＶ］もし集合Ａ上の頻度計数ＦがＦＣ内に、写像ｆ：Ａ→ＢがＭＡＰＳ内にあれば、f_*（Ｆ）がＦＣに加えられ得る（［ＦＣ　ＩＩ］参照）。このルールを使って頻度計数がＦＣに加えられるときは、ＦＣは使われた写像も記録する。
【００７４】
集合を頂点とし、写像を辺として、これらの集合は有向グラフ構造を形成すると考えられる。集合上の頻度計数もまた、頻度計数を頂点とし、写像を辺として、有向グラフ構造を形成すると考えられる。
【００７５】
これらの写像とデータは様々な順番でこれらのデータ構造に加えることができる。例えば、上記の木構造内で幅優先探索の順序を使うことができる。本実施例では、確率的アルゴリズムが使われる：
【００７６】
探索アルゴリズム
【００７７】
概要
【００７８】
以下の１から６の動作の何れかを確率的に実行せよ：
１．ＦＣ内の頻度計数ＦとＧの組を選んでＦ×ＧをＦＣに加える。Ｆが集合Ａ上、Ｇが集合Ｂ上であるとして、Ａ×ＢをＳＥＴＳに加える。
２．［Ｄ　ＩＩ］に従って写像に適用可能なＭＡＰＳ内の写像を選んで適用し、結果をＭＡＰＳに加える。
３．ＳＥＴＳ内の集合Ａを選び、Ａのある真部分集合ＢをＳＥＴＳに加え、$subset_B$：Ａ→ｂｏｏｌをＭＡＰＳに加える。
４．ＦＣ内の頻度計数Ｆを選ぶ。Ｆが集合Ａ上であるとして、ＳＥＴＳ内のＡの真部分集合Ｂを選び、$F|_B$をＦＣに加える。
５．ＭＡＰＳ内の写像ｆ：Ａ→Ｂを選び、ＳＥＴＳ内のＢの真部分集合Ｃを選ぶ。逆像f^{-1}（Ｃ）をＳＥＴＳに加える。
６．ＦＣ内の頻度計数Ｆと、Ｆが上にある集合から何か他の集合への写像ｆをＭＡＰＳ内に選び、f_*（Ｆ）をＦＣに加える。
【００７９】

詳細
【００８０】
図２は探索アルゴリズムのフローチャートを示す。動作とその対象の選択は確率的になされる。
【００８１】
【０１０４】
原始写像
【０１０５】
一般的な原始写像に加えて、画像に特に有用な原始写像を加えることもある。例えば、画像が普通そうであるように画素からなれば、画素間の隣接関係が有用であるかもしれない。これはＤｏｍの２要素が隣接画素であるときのみｔｒｕｅを返す原始写像
Ｎｂ：Ｄｏｍ×Ｄｏｍ→ｂｏｏｌとしてシステムに入れることができる。もう一つの例は、例えばウェーブレットフィルターなどの、画像処理の関連分野で知られる種々のフィルターである。
【０１０６】
誘導データと写像
【０１０７】
本方法がＭＡＰＳとＦＣに加えそうな写像とデータの簡単な例を挙げる：
【０１０８】
Ａ．色頻度
１．Ａ１．［Ｄ　Ｉ］より、２つの頻度計数Ｄｏｍ→Ｃｏｌ上のＩｍとＤｏｍ上のＳｔ（Ｄｏｍ）に基づいて、（Ｄｏｍ→Ｃｏｌ）×Ｄｏｍ上の頻度計数Ｉｍ×Ｓｔ（Ｄｏｍ）がＦＣに加えられる。
２．Ａ２．［Ｄ　ＩＶ］より、Ａ１からのＩｍ×Ｓｔ（Ｄｏｍ）と（原始写像なのでＭＡＰＳ内にある）評価写像ｅｖ：（Ｄｏｍ→Ｃｏｌ）×Ｄｏｍ→Ｃｏｌに基づいて、ev_*（Ｉｍ×Ｓｔ（Ｄｏｍ））がＦＣに加えられる。Ｃｏｌ上の頻度計数ev_*（Ｉｍ×Ｓｔ（Ｄｏｍ））は粒子（ｃ，n_c）の集合で、n_cは色ｃを持つ画素の数である。
【０１０９】
Ｂ．色の違いと位置の違いの頻度
１．Ｂ１．［Ｄ　ＩＩ］により、対角写像ｄｉａｇ：（Ｄｏｍ→Ｃｏｌ）→（Ｄｏｍ→Ｃｏｌ）×（Ｄｏｍ→Ｃｏｌ）、積写像ｍｐ：（Ｄｏｍ→Ｃｏｌ）×（Ｄｏｍ→Ｃｏｌ）→（Ｄｏｍ×Ｄｏｍ→Ｃｏｌ×Ｃｏｌ）、及び対角写像
ｄｉａｇ：Ｄｏｍ×Ｄｏｍ→（Ｄｏｍ×Ｄｏｍ）×（Ｄｏｍ×Ｄｏｍ）に基づいて、写像（$mp^{o}diag$）×ｄｉａｇ：（Ｄｏｍ→Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）→（Ｄｏｍ×Ｄｏｍ→Ｃｏｌ×Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）×（Ｄｏｍ×Ｄｏｍ）がＭＡＰＳに加えられる。
２．Ｂ２．［Ｄ　ＩＩ］より、評価写像ｅｖ：（Ｄｏｍ×Ｄｏｍ→Ｃｏｌ×Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）→Ｃｏｌ×Ｃｏｌ、及びＤｏｍ×Ｄｏｍの恒等写像に基づいて、写像ｅｖ×$id_{Dom\times Dom}$：（Ｄｏｍ×Ｄｏｍ→Ｃｏｌ×Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）×（Ｄｏｍ×Ｄｏｍ）→（Ｃｏｌ×Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）がＭＡＰＳに加えられる。
３．Ｂ３．［Ｄ　ＩＩ］より、色空間の引算と画像領域の差写像に基づいて、写像$Sub_{Col}\times Diff_{Dom}$：（Ｃｏｌ×Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）→Ｃｏｌ×V_{Dom}がＭＡＰＳに加えられる。
４．Ｂ４．［Ｄ　ＩＩ］によりＢ１、Ｂ２、Ｂ３でＭＡＰＳに加えられた写像を合成して（$Sub_{Col}\times Difr_{Dom}$）^{o}（ｅｖ×$id_{Dom\times Dom}$）^{o}（（$mp^{o}diag$）×ｄｉａｇ）：（Ｄｏｍ→Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）→Ｃｏｌ×V_{Dom}がＭＡＰＳに加えられる。
５．Ｂ５．［Ｄ　Ｉ］により、（Ｄｏｍ→Ｃｏｌ）×（Ｄｏｍ×Ｄｏｍ）上の頻度計数Ｉｍ×Ｓｔ（Ｄｏｍ×Ｄｏｍ）がＦＣに加えられる。

６．Ｂ６．［Ｄ　ＩＶ］により、Ｂ４の写像をＢ５で加えられた頻度計数Ｉｍ×Ｓｔ（Ｄｏｍ×Ｄｏｍ）に適用した結果がＦＣに加えられる。
Ｂ６で加えられたＣｏｌ×V_{Dom}上の頻度計数は粒子（（ｄ，ν），$n_{d,\nu}$）の集合で、$n_{d,\nu}$はｉ）色の違いｄを持ち、ｉｉ）それらの間の画像領域内でのベクトルがνである、画素の組の数である。

【０１１０】

パターン

【０１１１】

Ａ２で得られるＣｏｌ上の頻度計数ev_{*}（Ｉｍ×Ｓｔ（Ｄｏｍ））は、あまり多くの色が使われていないときに小さいエントロピーを持つ。画像全体が一色なら、エントロピーとして可能な最小値０を持つ。

【０１１２】

Ｂ６で加えられるＣｏｌ×V_{Dom}上の頻度計数は、同じ特定の色の違いと同じベクトルで隔てられたがその組がたくさんあるとき、小さいエントロピーを持つ。例えばもし一つの色の水平な直線があれば、色の違い０と水平ベクトルの粒子に比較的高い集中（計数の高い粒子）があり、この頻度計数のエントロピーは低くなる。

例２：データマトリクス

【０１１３】

データマトリクスとはＮ行Ｄ列の直方配列で、各行が異なる観察あるいは固体を与え、各列が異なる属性や変数を与えるものである。各変数は、ここで値集合と呼ぶ何かの集合の要素である値をとることができる。例えば、もし変数が制すうちだけをとり得るなら、値集合は整数の集合である。もし変数が任意の数をとり得るなら、値集合は実数の集合である。あるいは変数が「ｙｅｓ」か「ｎｏ」の値しかとれないなら、値集合はブール値の集合でありうる。

【０１１４】

Ｄ個の変数をa_1，a_2，．．．，a_Dで表し、それらの変数が値をとり得る集合をそれぞれX_1，X_2，．．．，X_Dで表す。すると、各観察は集合$X_1 \times X_2 \times \ldots \times X_D$の要素を与える。データマトリクスの形をとった入力データは、本実施例では各観察が一つの粒子の１計数に寄与する$X_1 \times X_2 \times \ldots \times X_D$上の頻度計数として表される。従って頻度計数のマスはＮである。

【産業上の利用可能性】

【０１１５】

以上のように、高次元データがより効果的に解析され得るように、また与えられたデータ内によりよくパターンを見つけられるように、与えられたデータを配置する方法と装置が開示された。本発明は広い範囲の産業で利用可能である。それらの産業では、ますます多くのデータが収集され、巨大なデータの山から適切な情報を見つけ出すことがますます重要になっている。本発明が有用な分野は、多数の遺伝子と、ある遺伝病に罹患する比較的少数の患者の例、及び用意に百万次元（画素）を持ちうる画像の例を含む。

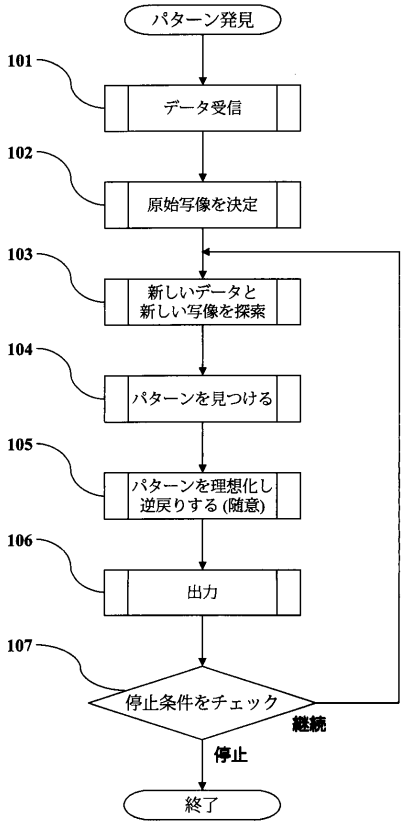
【０１１６】

本明細書には本発明の特定の好ましい特徴のみを説明記述したが、関連技術分野で熟練した者には多くの修正や改変が思い浮かぶであろう。例えば、本発明を説明するためにここで使われた集合や写像の概念は、様々な分野で多くの同値あるいは類似の概念を持つ。例えば、関数、型、メソッド等である。集合や写像などの用語は、望むならば完全に避けることが可能である。本発明全体をデータとサブルーチンの言葉で記述することも可能である。しかし、そのような表面的な違いは、真の違いではない。

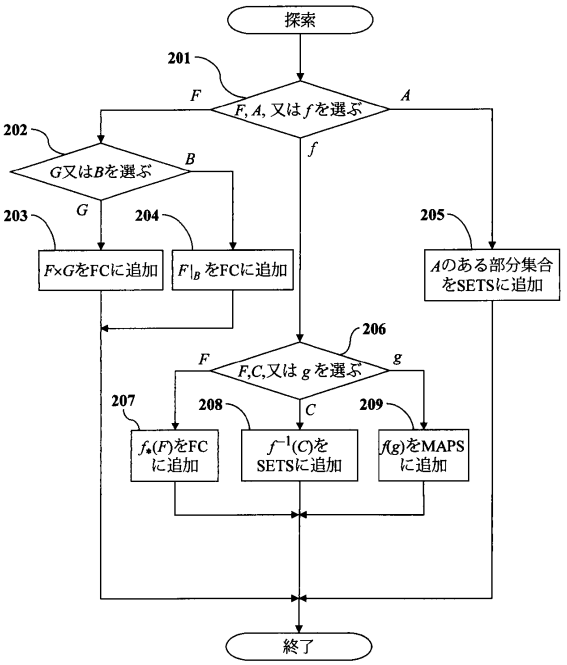
【０１１７】

従って、添付の特許請求の範囲は、そのような全ての修正、変更、用語の違いを本発明の真の精神のうちに入るものとして全て含むよう意図されたものであることを理解されたい。

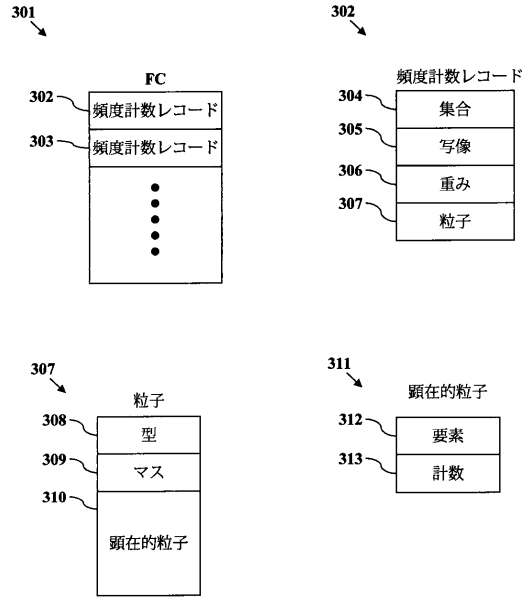
【図１】


パターン発見
101
データ受信
102
原始写像を決定
103
新しいデータと
新しい写像を探索
104
パターンを見つける
105
パターンを理想化し
逆戻りする (随意)
106
出力
107
停止条件をチェック
継続
停止
終了

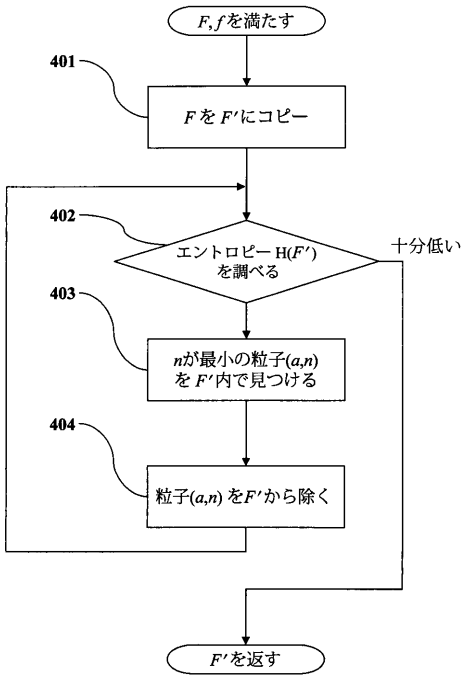
【図２】


探索
201
F, A, 又は f を選ぶ
F
A
f
202
G又はBを選ぶ
B
G
203
F×GをFCに追加
204
F|B をFCに追加
205
Aのある部分集合
をSETSに追加
206
F,C,又は g を選ぶ
F
g
C
207
f*(F)をFC
に追加
208
f⁻¹(C)を
SETSに追加
209
f(g)をMAPS
に追加
終了

【図３】


301
FC
302
頻度計数レコード
303
頻度計数レコード
302
頻度計数レコード
304
集合
305
写像
306
重み
307
粒子
307
粒子
308
型
309
マス
310
顕在的粒子
311
顕在的粒子
312
要素
313
計数

【図４】


F, f を満たす
401
F を F′にコピー
402
エントロピー H(F′)
を調べる
十分低い
403
nが最小の粒子(a,n)
を F′内で見つける
404
粒子(a,n) をF′から除く
F′を返す