

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6001871号
(P6001871)

(45) 発行日 平成28年10月5日 (2016. 10. 5)

(24) 登録日 平成28年9月9日 (2016. 9. 9)

(51) Int. Cl.

F I

G 0 6 F 19/00 (2011.01)

G 0 6 F 19/00 1 3 0

請求項の数 10 外国語出願 (全 11 頁)

(21) 出願番号 特願2012-28991 (P2012-28991)
 (22) 出願日 平成24年2月14日 (2012. 2. 14)
 (65) 公開番号 特開2012-168949 (P2012-168949A)
 (43) 公開日 平成24年9月6日 (2012. 9. 6)
 審査請求日 平成27年2月2日 (2015. 2. 2)
 (31) 優先権主張番号 13/027, 829
 (32) 優先日 平成23年2月15日 (2011. 2. 15)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 390041542
 ゼネラル・エレクトリック・カンパニイ
 アメリカ合衆国、ニューヨーク州 1 2 3
 4 5、スケネクタデイ、リバーロード、1
 番
 (74) 代理人 100137545
 弁理士 荒川 聡志
 (74) 代理人 100105588
 弁理士 小倉 博
 (74) 代理人 100129779
 弁理士 黒川 俊久
 (74) 代理人 100113974
 弁理士 田中 拓人

最終頁に続く

(54) 【発明の名称】 混合モデルの構築方法

(57) 【特許請求の範囲】

【請求項 1】

非一時的メディア (1 0 2) に格納されるデータセットの一般的な混合モデル (1 0 0) を生成する方法であって、

前記データセットの 1 つ以上の次元の 1 つ以上の要素を含み、前記データセットのサブセットを定義するためのサブセット基準 (1 0 8) を提供するステップと、

前記サブセット基準 (1 0 8) を利用して、プロセッサ上で前記データセットを通じてフィルタリングするステップと、

前記プロセッサ内で前記サブセット基準 (1 0 8) に基づいて前記データセットをサブセットに分割するステップであって、前記サブセットの数は前記サブセット基準 (1 0 8) 内の要素の数と等しい、ステップと、

各サブセットに識別子を割り当てるステップと、

前記サブセットのそれぞれについてサブセット混合モデル (1 1 0) を生成するステップと、

少なくとも 2 つのサブセットがある場合には、前記少なくとも 2 つのサブセットのそれぞれの前記サブセット混合モデルを前記一般的な混合モデル (1 2 0) に結合し、少なくとも 2 つのサブセットがない場合には、前記サブセット混合モデルを前記一般的な混合モデル (1 2 0) として設定するステップと、

を含む、方法。

【請求項 2】

前記サブセット基準が、関係データベースにおいて定義されるステップと、少なくとも1つの次元によって前記データセットをフィルタリングするステップとのうちの1つを含む、請求項1に記載の方法。

【請求項3】

前記生成するステップが、サブセット(104)の少なくとも1つのコンポーネントを識別するステップと、関数をサブセットの少なくとも1つのコンポーネントに適合させるステップと、スケーリングファクタによってフィッティング関数をスケーリングするステップと、スケーリングされたフィッティング関数を合計するステップとのうちの少なくとも1つを含む、請求項1または2に記載の方法。

【請求項4】

前記関数が確率密度関数である、請求項3に記載の方法。

【請求項5】

前記確率密度関数が正規分布関数である、請求項4に記載の方法。

【請求項6】

前記スケーリングファクタがスカラー値である、請求項3に記載の方法。

【請求項7】

サブセットの前記フィッティング関数のそれぞれに対応する前記スケーリングファクタの全ての合計が1である、請求項4に記載の方法。

【請求項8】

前記結合するステップ(120)が、前記少なくとも1つのサブセットごとに前記サブセット混合モデルを連結するステップと、前記少なくとも1つのサブセットごとに前記サブセット混合モデルを別々にスケーリングして前記スケーリングされたサブセット混合モデルを連結するステップと、前記サブセット混合モデル(150)を結合する前に1つまたは複数のコンポーネント関数を除去するステップとを備える、請求項1または2に記載の方法。

【請求項9】

前記サブセット混合モデルを結合する前に1つまたは複数のコンポーネント関数を除去するステップ(150)が、コンポーネントを選択して、前記選択されたコンポーネントと前記選択されたコンポーネントに対応する前記サブセット以外のサブセットからの全ての前記コンポーネントとの間の前記距離を決定するステップ(144)を備える、請求項8に記載の方法。

【請求項10】

前記サブセット混合モデルを結合する前に1つまたは複数のコンポーネント関数を前記除去するステップ(150)が、最大距離を有する前記コンポーネントを除去するステップをさらに備える、請求項9に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本願は、混合モデルの構築方法に関する。

【背景技術】

【0002】

データマイニングとは、データから情報および値を抽出するために使用される技術である。データマイニングアルゴリズムは、ターゲットマーケティングのための買い物の消費傾向の予測、クレジットカード不正取引の検出、ウェブサイト内の顧客のナビゲーションパスの予測、機械の故障検出などの、多くのアプリケーションで使用されている。データマイニングは、人工知能(AI)および統計モデリングコミュニティによって長年にわたって開発されてきた様々なアルゴリズムを使用する。アルゴリズムには異なる多くのクラスがあるが、それらは全て(a)データドメインの知識を表す(暗黙的に、または明示的に)モデル、(b)モデルを構築するために訓練データを使用するモデル構成か学習フ

10

20

30

40

50

エーズ、および(c)予測を行うために新しいデータを採用して、そのデータにモデルを適用する推論ファシリティなどの、いくつかの共通の特徴を共有する。知られている例は、第2変数の値に重み付けして、加重値と一定値とを合計することによって、第2変数から第1変数が予測される、線形回帰モデルである。重みおよび一定値はモデルのパラメータである。

【0003】

混合モデルは、G McLachlan and D Peel in Finite Mixture Models, John Wiley & Sons, (2000)によって記述されるように、学術研究コミュニティ内のデータマイニングアプリケーションのために一般に使用されるモデルである。混合モデルのクラスには、混合エキスパートモデル(Mixtures of Experts)および階層型混合エキスパートモデル(Hierarchical Mixtures of Experts)などのバリエーションがある。文書により十分に立証された、混合モデルを構築するためのアルゴリズムもある。一例は、期待値最大化(EM)である。このような混合モデルは、一般に、データ内のクラスタまたはコンポーネントを識別すること、およびそれぞれのクラスタに適切な数学関数を適合させることによって構築される。

【発明の概要】

【0004】

一態様では、非一時的メディアに格納されるデータセットの一般的な混合モデルを生成する方法は、データセットのサブセットを定義するためのサブセット基準を提供するステップと、プロセッサ内でサブセット基準に基づいてデータセットを少なくとも2つのサブセットに分割するステップと、少なくとも2つのサブセットのそれぞれについてサブセット混合モデルを生成するステップと、少なくとも2つのサブセットのそれぞれのサブセット混合モデルを一般的な混合モデルに結合するステップとを備える。

【図面の簡単な説明】

【0005】

【図1】本発明の一実施形態による一般的な混合モデル生成する方法を示す流れ図である。

【図2】図1で示される方法の一部として、サブセット混合モデルからコンポーネントをフィルタリングする方法を示す流れ図である。

【図3】図1の一般的な混合モデルを生成する方法による、データセットのフィルタリングの例を示す図である。

【図4】第1サブセットのサブセット混合モデルを示す図である。

【図5】第2サブセットのサブセット混合モデルを示す図である。

【図6】図1で開示された方法によって構築された一般的な混合モデルを示す図である。

【発明を実施するための形態】

【0006】

以下の記述では、説明の目的で、本明細書に開示される技術の完全な理解を提供するために特定の多くの詳細を説明する。しかし、これらの特定の詳細がなくても例示的实施形態を実施できることが当業者には明らかであろう。他の例では、例示的实施形態の説明を容易にするために、構造および装置が図の形式で示される。

【0007】

以下で、図面を参照して例示的实施形態を記述する。これらの図面は、本明細書で記述されるモジュール、方法、およびコンピュータプログラム製品を実装する特定の实施形態のいくつかの詳細を示している。しかし、これらの図面は、図面内に存在する場合がある何らかの制限を課すものと見なされるべきではない。方法およびコンピュータプログラム製品は、それらの動作を遂行するために、どのような機械可読メディア上に提供されてもよい。实施形態は、既存のコンピュータプロセッサを使用して、あるいはこのまたは他の目的のために組み込まれた専用コンピュータプロセッサによって、あるいはハードワイヤードシステムによって実装され得る。

【 0 0 0 8 】

上述のように、本明細書に記述される実施形態は、機械実行可能命令またはデータ構造を担持するまたは格納した機械可読メディアを備えるコンピュータプログラム製品を含む。このような機械可読メディアは、汎用または専用コンピュータ、あるいはプロセッサを備えた他の機械によってアクセスできる、どのような利用可能なメディアでもよい。例を挙げると、このような機械可読メディアは、RAM、ROM、EPROM、EEPROM、CD-ROM、あるいは他の光ディスク記憶装置、磁気ディスク記憶装置、または他の磁気記憶装置、あるいは所望のプログラムコードを機械実行可能命令またはデータ構造の形式で搬送または格納するために使用でき、汎用または専用コンピュータあるいはプロセッサを備えた他の機械によってアクセスできる、他の何らかのメディアを備えることができる。ネットワークまたは他の通信接続（ハードワイヤード、無線、またはハードワイヤードと無線の組合せでもよい）を介して機械に情報が転送または提供される際、機械はその接続を適切に機械可読メディアと見なす。したがって、このようななどの接続も適切に機械可読メディアと呼ばれる。上記の組合せも、機械可読メディアの範囲に含まれる。機械可読命令は、たとえば、汎用コンピュータ、専用コンピュータ、または専用処理機械に、ある機能または機能のグループを実行させる命令およびデータを備える。

10

【 0 0 0 9 】

たとえばネットワーク化された環境における機械によって実行されるプログラムモジュールの形式で、プログラムコードなどの機械可読命令を含むプログラム製品によって一実施形態で実装され得る方法ステップの一般的な文脈において実施形態を記述する。一般にプログラムモジュールは、特定のタスクの実行または特定の抽象データタイプの実装の技術的効果を有するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。機械可読命令、関連データ構造、およびプログラムモジュールは、本明細書に開示される方法のステップを実行するためのプログラムコードの例を表す。このような実行可能命令または関連データ構造の特定のシーケンスは、このようなステップに記述される機能を実装するための対応する動作の例を表す。

20

【 0 0 1 0 】

実施形態は、プロセッサを有する1つまたは複数の遠隔コンピュータへの論理的な接続を使用してネットワーク化された環境において実施され得る。論理的接続は、限定するためではなく例示のために本明細書で提示される、ローカルエリアネットワーク（LAN）およびワイドエリアネットワーク（WAN）を含むことができる。このようなネットワーク環境はオフィス全体または企業全体のコンピュータネットワーク、イントラネット、およびインターネットにおいて普通のことであり、異なる様々な通信プロトコルを使用できる。このようなネットワークコンピューティング環境は、一般にパーソナルコンピュータ、ハンドヘルド装置、マルチプロセッサシステム、マイクロプロセッサを用いるまたはプログラム可能である家庭用電化製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、および同様のものを含む多くのタイプのコンピュータシステム構成を包含することが、当業者なら理解できるであろう。

30

【 0 0 1 1 】

実施形態は、通信ネットワークを通じてリンクされた（ハードワイヤードリンク、無線リンク、あるいはハードワイヤードまたは無線リンクの組合せによって）ローカルおよび遠隔処理装置によってタスクが実行される分散コンピューティング環境でも実施できる。分散コンピューティング環境では、プログラムモジュールはローカルメモリ記憶装置と遠隔メモリ記憶装置との両方に位置できる。

40

【 0 0 1 2 】

例示的实施形態の全体または一部を実装するための例示的システムは、処理装置、システムメモリ、およびシステムメモリを含む様々なシステムコンポーネントを処理装置に結合するシステムバスを含むコンピュータの形式の汎用コンピューティング装置を含むことができる。システムメモリは、読出し専用メモリ（ROM）およびランダムアクセスメモリ（RAM）を含むことができる。コンピュータは、磁気ハードディスクから読み取る、

50

またはそこに書き込むための磁気ハードディスクドライブ、リムーバブル磁気ディスクから読み取る、またはそこに書き込むための磁気ディスクドライブ、CD-ROMまたは他の光メディアなどのリムーバブル光ディスクから読み取る、またはそこに書き込むためのおよび光ディスクドライブも含むことができる。ドライブおよびその関連機械可読メディアは、機械可読命令、データ構造、プログラムモジュール、およびコンピュータのための他のデータの非揮発性記憶装置を提供する。

【0013】

実施形態で開示される方法の技術的効果は、予測パターンについての複雑なデータセットをマイニングするための正確なモデルをより効果的に提供することを含む。方法は、異なる問題を解決することを課せられた基本的に単一のアルゴリズムを使用して異なる観点からデータを探索するための高度の柔軟性を導入する。したがって、技術的効果はより効果的なデータ探索、異常検出、値を予測して欠損データを置換するための回帰、およびデータの区分化を含む。開示された方法を使用してこのようなデータを効果的に探索できる方法の例には、顧客の購買傾向に基づくターゲットマーケティング、リスクの高いクレジット申込者を識別することによるクレジットリスクの減少、および航空機の状態を理解することから行われる予測的なメンテナンスがある。

【0014】

本発明は、データセットの一般的な混合モデルを生成することに関する。より詳細には、データセットは2つ以上のサブセットに分割され、サブセットごとにサブセット混合モデルが生成され、次いでサブセット混合モデルが結合されてデータセットの一般的な混合モデルを生成する。

【0015】

図1を参照すると、一般的な混合モデル100を生成する方法が開示されている。まず、サブセット識別を有するサブセットを生成する(104)ために、データベース102内に含まれるデータセットとサブセット基準108が提供される。構成データセットを有するデータセットを電子メモリに格納できる。データセットは多次元またはパラメータを含むことができ、それぞれの次元はそれに関連付けられる1つまたは複数の値を有する。値は離散値でもよく、連続値でもよい。たとえば、データセットは、CFM56、CF6、CF34、GE90、およびGENxの離散値を有するガスタービンエンジンの次元を備えることができる。離散値は、General Electric Corporationによって製造および販売されるガスタービンエンジンの様々なモデルを表す。データセットは、B737-700、B737700ER、B747-8、B777-200LR、B777-300ER、およびB787の離散値を有する機体と呼ばれる別の次元をさらに備えることができ、データセットのガスタービンエンジン次元のガスタービンエンジンを搭載できる様々な機体を表している。この例を続けると、データセットは、18,000重量ポンドから115,000重量ポンド(80ノット~512ノット)の範囲内の値などの連続値を有するスラストと呼ばれる次元をさらに備えることができる。

【0016】

サブセット基準108は、データセットをフィルタリングするために使用できるデータセットの1つまたは複数の次元の1つまたは複数の値でよい。サブセット基準は、関係データベースに格納することもでき、他の何らかの知られている方法によって設計することもできる。一般に、サブセット基準108は、ユーザがデータセットから学びたいものに基づいて、データセットのユーザによって構築される。サブセット基準108は、データセット内のデータをフィルタリングおよび分割するためのいくつかの個々の基準を含むことができる。上記の例を続けると、サブセット基準108は、B747-8に搭載されたGE90エンジン、B777-300ERに搭載されたGENxエンジン、およびB787に搭載されたGENxなどの3つの異なる要素を備えることができる。これは3つの要素を備える2次元サブセット基準の例であるが、サブセット基準はデータセット内の次元の数を最大とするいくつかの次元を含むことができ、いくつかの要素を備えることができる。

【 0 0 1 7 】

サブセットおよびサブセット識別を生成するステップ (1 0 4) は、データセットを通じてフィルタリングするステップ、およびそれぞれのサブセット内のそれぞれの要素を識別するステップを備える。サブセットの数は選択基準内の要素の数と等しい。フィルタリング処理は、データベース 1 0 2 を備える電子メモリへのアクセスを有するプロセッサ上で実行しているコンピュータソフトウェア要素によって遂行できる。フィルタリングの後、またはそれと同時に、サブセットおよびその構成要素とそれぞれの他のサブセットおよびそれらの構成要素とを区別するために、それぞれのサブセットにサブセット識別子が割り当てられる。サブセット識別子はテキスト文字列でもよく、1 0 4 で生成されたサブセットを識別する他の知られている方法でもよい。

10

【 0 0 1 8 】

次に、1 0 6 で、少なくとも 1 つのサブセットがあるかどうか評価される。少なくとも 1 つのサブセットがない場合、方法 1 0 0 は 1 0 8 に戻って、少なくとも 1 つのサブセットを作成する新しいサブセット基準を受け取る。少なくとも 1 つのサブセットがある場合は、方法 1 0 0 は 1 1 0 でサブセットごとに混合モデルを生成する。混合モデルの生成は、一般にデータマイニングの分野における訓練とも呼ばれる。サブセットごとの混合モデルは、何らかの知られている方法によって、また何らかの知られているタイプの混合モデルとして生成でき、非限定的な例は、期待値最大化 (E M) を使用して訓練されたガウス混合モデルである。サブセットごとに混合モデルを生成する処理は、サブセット密度を表す数学的汎関数をもたす。連続ランダムベクトルをモデリングする例では、それぞれのサブセットの数学的汎関数の表現は、確率密度関数 (p d f) のスケーリングされた (s c a l e d) 合計である。それぞれの p d f は、混合モデルが生成されているサブセット内のデータ要素のコンポーネントまたはクラスタに対応する。言い換えれば、それぞれのサブセット 1 1 0 の混合モデルを生成する方法は、プロセッサ上で実行しているソフトウェア要素によって行われ、ソフトウェア要素はサブセット内の全てのデータ要素を考慮し、データ要素を 1 つまたは複数のコンポーネントにクラスタリングし、それぞれのコンポーネントに p d f を適合させ、それぞれのコンポーネントにスケーリングファクタを帰着して (a s c r i b e) 、データの数学的汎関数の表現を生成する。混合モデルの非限定的な例は、

20

【 0 0 1 9 】

30

【 数 1 】

$$p(X) = \sum_{k=1}^K \pi_k N(X | \mu_k, \Sigma_k)$$

の形式のガウスまたは正規分布混合モデルであり、

上式で、 $p(x)$ はサブセットの数学的汎関数の表現であり、

X は変数の多次元ベクトル表現であり、

k はサブセット内のそれぞれのコンポーネントを指すインデックスであり、

K はサブセット内のコンポーネントの総数であり、

π_k は、1 と等しい全ての K クラスタについての全ての π_k の合計を有するクラスタ k に対応するスカラー スケーリングファクタであり、

40

【 0 0 2 0 】

【 数 2 】

$$N(X | \mu_k, \Sigma_k)$$

は、コンポーネント平均 (c o m p o n e n t m e a n) μ_k および共分散 Σ_k についてのベクトル X の正規確率密度関数である。

【 0 0 2 1 】

ベクトル X が一次元のベクトルの場合 Σ_k は X の分散であり、 X が 2 つ以上の次元を有する場合 Σ_k は X の共分散行列である。

50

【 0 0 2 2 】

1 1 0でサブセットごとに混合モデルが生成された後、1 1 2で、少なくとも2つのサブセットがあるかどうか決定される。少なくとも2つのサブセットがない場合、1 1 0で生成された単一のサブセット混合モデルが一般的な混合モデルである。しかし、1 1 2で少なくとも2つのサブセットがあると決定される場合、次に1 1 6でモデルコンポーネントのフィルタリングが所望されるかどうか決定される。1 1 6でフィルタリングが所望される場合、1 1 8で、1つまたは複数のコンポーネントがモデルから除去される。1 1 8のフィルタリング方法は、図2とともにより詳細に記述する。一旦1 1 8でフィルタリングが行われると、または1 1 6でフィルタリングが所望されない場合、方法1 0 0は1 2 0に進み、そこでサブセットモデルが結合される。

10

【 0 0 2 3 】

1 2 0でサブセットモデルを結合するステップは、サブセットごとに生成された混合モデルを連結して結合モデルを生成するステップを備えることができる。あるいは、サブセットモデルを結合するステップは、それぞれの混合モデルを連結して結合モデルを生成する前に、個々のサブセットのそれぞれの混合モデルを別々にスケーリングするステップを備えることができる。

【 0 0 2 4 】

1 2 2で、モデルの単純化が所望されるかどうか決定される。1 2 2で単純化が所望されない場合、1 2 4で、結合されたサブセットモデルが一般的なモデルである。1 2 2で単純化が所望される場合、結合モデルの単純化が1 2 6で実行され、1 2 8で、単純化された結合モデルが一般的なモデルと見なされる。単純化1 2 6は、2つ以上の異なるサブセットから1つまたは複数のクラスタを結合するステップを備えることができる。単純化1 2 6は、サブセットの結合された混合モデルから1つまたは複数のコンポーネントを除去するステップをさらに備えることができる。

20

【 0 0 2 5 】

次に図2を参照すると、サブセット混合モデルを結合する前に、1 1 8で個々のサブセット混合モデルのコンポーネントをフィルタリングする方法が示されている。まず、1 4 0でそれぞれのコンポーネントおよび他のコンポーネントへの関連距離を一覧にするための完成リストが消去される。次に、1 4 2で、全てのサブセットから全てのコンポーネントが、プロセッサおよび関連電子メモリによって受信される。1 4 4で全てのコンポーネントからあるコンポーネントが選択され、1 4 6で、選択されたコンポーネントの、他のサブセット内の他の全てのコンポーネントへの距離が決定される。言い換えれば、選択されたコンポーネントが、選択されたコンポーネントのサブセット識別子とは異なるサブセット識別子を有する他の全てのコンポーネントと比較される。距離は、これに限定されないが、カルバックライブラー情報量を含む、何らかの知られている方法によって計算できる。1 4 8で、コンポーネントおよび他のサブセットの他の全てのコンポーネントへの関連距離が一覧にされて、完成リストに添付される。言い換えれば、完成リストはコンポーネントから他のサブセットの全てのコンポーネントへの距離を含む。1 5 0で、選択されたコンポーネントが最後のコンポーネントかどうか決定される。最後のコンポーネントではない場合、方法1 1 8は1 4 4に戻って次のコンポーネントを選択する。しかし、1 5 0で、選択されたコンポーネントが最後のコンポーネントと決定される場合、全てのサブセットの全てのコンポーネントについて完成リストがアップデートされ、方法は1 5 2に進み、完成リストが1 4 6で計算された距離の降順でソートされる。1 5 4で、完成リストのトップのコンポーネント、または他の全てのサブセットの他の全てのコンポーネントへの最大距離を有するコンポーネントが除去されるかフィルタリングによって取り除かれる。1 5 6で、フィルタリング基準が満たされているかどうか決定される。フィルタリング基準は、たとえば、あらかじめ定められたフィルタリングされるべきコンポーネントの総数でよい。あるいは、フィルタリング基準は、コンポーネントの総数のあらかじめ定められた割合のフィルタリングでよい。1 5 6でフィルタリング基準が満たされる場合、1 6 0で、最後のコンポーネントセットが識別される。しかし、1 5 6でフィルタリング基

30

40

50

準が満たされない場合、158で、反復フィルタリングが所望されるかどうか決定される。反復フィルタリングへの所望は方法118のユーザによって設定される。158で反復フィルタリングが所望されない場合、方法は154に戻って、他のサブセットから他の全てのコンポーネントへの最大距離を有するコンポーネントを残りのコンポーネントから除去する。158で、反復フィルタリングが所望されると決定されると、方法118は140に戻る。

【0026】

反復フィルタリングは、方法118がコンポーネントごとにそれぞれの他のコンポーネントへの距離を再計算して、混合モデルからコンポーネントが除去されるたびに140から152を実行することによって新しい完成リストを生成することを意味する。コンポーネント間の距離は変更でき、したがって、混合モデルからコンポーネントが除去されると完成リスト上のコンポーネントの相対的順序は変更できる。したがって、反復フィルタリングを実行することによって、除去されるコンポーネントがそれぞれの他のサブセットからコンポーネントへの最大距離を有するコンポーネントであることを、より自信を持って確実にすることができる。しかし、場合によっては、反復フィルタリングはより計算集約型であってより時間がかかるので、反復フィルタリングを実行したくない場合がある。言い換えれば、本明細書に開示されたフィルタリング方法118を実行すると、158で反復フィルタリングが所望されるかどうか決定するために、フィルタリング実行とフィルタリングに必要な時間と間のトレードオフを評価できる。

【0027】

図3～6は、一般的な混合モデルを生成する前述の方法100を実行する例を示している。図3で、データセットからのデータ180および190が、変数 x_1 に対してプロットされている。データは、方法100の104とともに記述される手順に従って、グラフ上に白丸で示されている第1サブセット180、およびグラフ上に黒い三角形で示されている第2サブセット190にさらに分割される。方法100を多くのサブセットを有する多変量解析に適用できるが、この例では方法100の視覚化を単純にするために、2つのサブセットだけに依存する単一の変数データが示される。

【0028】

図4および5は、ステップ110として、それぞれ第1サブセット180のための、および第2サブセット190のための混合モデルの生成を示している。第1サブセット180の場合、3つのコンポーネントが識別され、それぞれに平均 μ_1 、 μ_2 、および μ_3 を有するスケールされたガウス分布 G_1 、 G_2 、および G_3 がそれぞれ適用される。第2サブセット190の場合、2つのコンポーネントが識別され、それぞれに平均 μ_4 および μ_5 を有するスケールされたガウス分布 G_4 および G_5 がそれぞれ適用される。したがって、第1サブセット180の混合モデルは、構成コンポーネント G_1 、 G_2 、および G_3 のスケールされたフィッティング関数のエンベロープによって表される。同様に、第2サブセット190の混合モデルは、構成コンポーネント G_4 および G_5 のスケールされたフィッティング関数のエンベロープによって表される。図6で、フィルタリング後、方法100のステップ120として、一般的な混合モデルの結合された構成スケールフィッティング関数が示されている。この例では、フィルタリングステップ118で、フィッティング関数 G_3 を有するコンポーネントは、いくつかのあらかじめ定められた値(図示せず)を超える他のサブセット G_4 および G_5 のコンポーネントから離れていることが見付き、したがってコンポーネント G_3 は図6の一般的な混合モデルから除去されたことがわかる。

【0029】

本明細書は本発明を開示するために、およびあらゆる当業者が本発明を利用および使用できるようにするために、ベストモードを含めた例を用いる。本発明の特許性の範囲は特許請求の範囲によって定義され、当業者が想到する他の例を含むことができる。このような他の例は、それらの例が特許請求の範囲の文言とは異なる構造的要素を有する場合、またはそれらの例が特許請求の範囲の文言とは実質的に差のない均等の構造要素を含む

10

20

30

40

50

場合、特許請求の範囲内であるものとする。

【符号の説明】

【0030】

100	一般的な混合モデルを生成する方法	
102	データベース	
104	サブセット識別を有するサブセットを生成する	
106	少なくとも1つのサブセットがあるか？	
108	サブセット基準	
110	サブセットごとに混合モデルを構築する	
112	少なくとも2つのサブセットがあるか？	10
114	サブセットモデルが一般的なモデルである	
116	モデルフィルタリングが所望されるか？	
118	モデルからコンポーネントを除去する	
120	サブセットモデルを結合する	
122	単純化が所望されるか？	
124	結合されたサブセットモデルが一般的なモデルである	
126	結合されたモデルを単純化する	
128	単純化されたモデルが一般的なモデルである	
140	完成リストを初期化して空にする	
142	全てのサブセットから全てのコンポーネントを受信する	20
144	次のコンポーネントを選択する	
146	選択されたコンポーネントの、他のサブセット内の他の全てのコンポーネント	
への距離を計算する		
148	選択されたコンポーネントおよび距離が完成リストに添付される	
150	選択されたコンポーネントは最後のコンポーネントか？	
152	完成リストを降順でソートする	
154	トップのコンポーネントを除去する	
156	フィルタリング基準は満たされたか？	
158	反復フィルタリングが所望されるか？	
160	最後のコンポーネントセットが識別される	30
180	第1サブセットデータ	
190	第2サブセットデータ	
G1	第1クラスタフィッティング関数	
G2	第2クラスタフィッティング関数	
G3	第3クラスタフィッティング関数	
G4	第4クラスタフィッティング関数	
G5	第5クラスタフィッティング関数	
μ 1	第1クラスタの平均	
μ 2	第2クラスタの平均	
μ 3	第3クラスタの平均	40
μ 4	第4クラスタの平均	
μ 5	第5クラスタの平均	

【図 1】

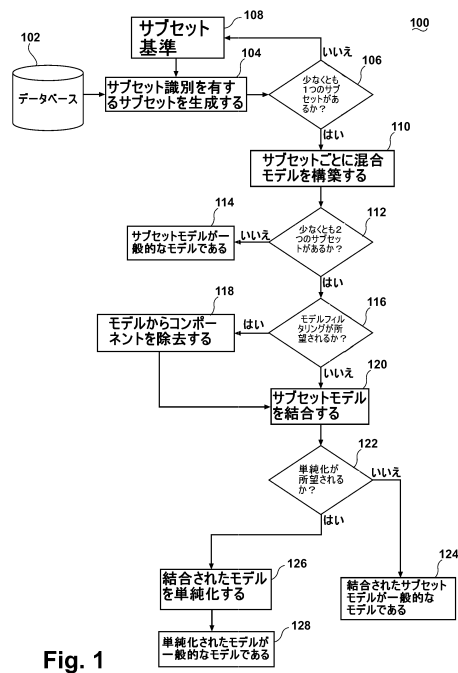


Fig. 1

【図 2】

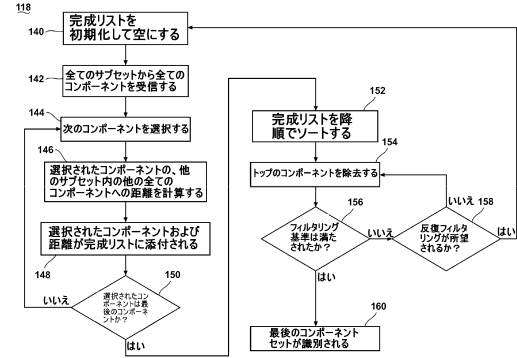


Fig. 2

【図 3】

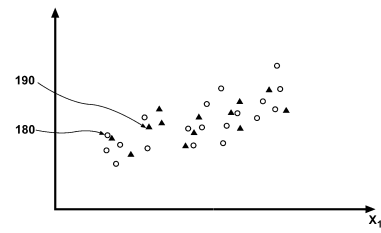


Fig. 3

【図 4】

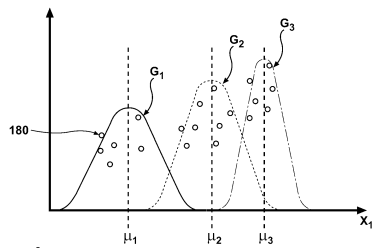


Fig. 4

【図 6】

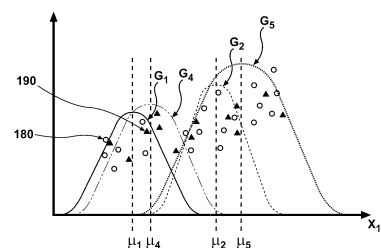


Fig. 6

【図 5】

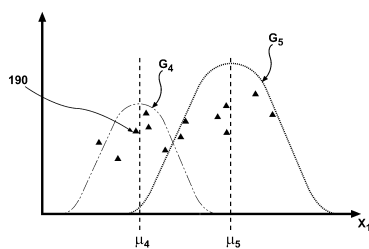


Fig. 5

フロントページの続き

- (72)発明者 ロバート・エドワード・カラン
イギリス、ハンプシャー・エスオー53・4ワイジー、イーストレイ、スクール・レーン、チャン
ドラーズ・フォード、ジーイー・アピエーション・システムズ・リミテッド
- (72)発明者 ブライアン・ラーデル
イギリス、ハンプシャー・エスオー53・4ワイジー、イーストレイ、スクール・レーン、チャン
ドラーズ・フォード、ジーイー・アピエーション・システムズ・リミテッド

審査官 梅岡 信幸

- (56)参考文献 村松 正吾，標準の記法を使って仕様をもむ，バグを出す，コードを作る 組み込みシステム開
発に“損して得とる”モデリング，Interface，日本，CQ出版株式会社，2007年
10月1日，第33巻，第10号，p.52-65
- ビショップ C.M.，パターン認識と機械学習 上 ベイズ理論による統計的予測，日本，シ
ュプリンガー・ジャパン株式会社，2007年12月10日，p.107-110

(58)調査した分野(Int.Cl.，DB名)

G06Q 10/00 - 99/00
G06F 19/00