

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5425814号  
(P5425814)

(45) 発行日 平成26年2月26日(2014.2.26)

(24) 登録日 平成25年12月6日(2013.12.6)

(51) Int.Cl.

F I

GO 1 N 33/48 (2006.01)

GO 1 N 33/48 M

GO 1 N 33/49 (2006.01)

GO 1 N 33/49 Y

GO 1 N 21/17 (2006.01)

GO 1 N 33/49 K

GO 1 N 15/14 (2006.01)

GO 1 N 21/17 A

GO 1 N 21/64 (2006.01)

GO 1 N 15/14 Z

請求項の数 18 (全 14 頁) 最終頁に続く

(21) 出願番号 特願2010-546084 (P2010-546084)  
 (86) (22) 出願日 平成21年2月8日(2009.2.8)  
 (65) 公表番号 特表2011-515655 (P2011-515655A)  
 (43) 公表日 平成23年5月19日(2011.5.19)  
 (86) 国際出願番号 PCT/US2009/033504  
 (87) 国際公開番号 W02009/100410  
 (87) 国際公開日 平成21年8月13日(2009.8.13)  
 審査請求日 平成24年2月8日(2012.2.8)  
 (31) 優先権主張番号 61/027, 416  
 (32) 優先日 平成20年2月8日(2008.2.8)  
 (33) 優先権主張国 米国 (US)

(73) 特許権者 510215581  
 ヘルス ディスカバリー コーポレイショ  
 ン  
 HEALTH DISCOVERY CO  
 RPORATION  
 アメリカ合衆国 ジョージア州 3140  
 1, サバナ, スイート601, イーストブ  
 ライアンストリート 2  
 (74) 代理人 100096024  
 弁理士 柏原 三枝子  
 (74) 代理人 100125520  
 弁理士 高橋 剛一  
 (74) 代理人 100155310  
 弁理士 柴田 雅仁

最終頁に続く

(54) 【発明の名称】 サポートベクタマシンを用いてフローサイトメトリーデータを分析するための方法及びシステム

## (57) 【特許請求の範囲】

## 【請求項 1】

フローサイトメトリーデータを分析及び分類するための方法であって、前記方法が：

フローサイトメトリー分析によって複数のサンプルについて実施される複数の分析試料を  
 具える入力データセットを、プロセッサ及び記憶装置を具えるコンピュータシステムの中  
 にダウンロードするステップを具えており、

前記プロセッサが、少なくとも1のサポートベクタマシンを実行するようプログラムさ  
 れ、

1又はそれ以上の特徴選択アルゴリズムを具える前記入力データセットの訓練データセ  
 ットを予処理して、前記複数の分析試料から減少した数の分析試料を選択するステップと 10

；  
 分散カーネルを具えるサポートベクタマシンを訓練して、前記減少した数の分析試料に  
 対応する前記訓練データセットを2つの確率分布に分け、前記2つの確率分布間の相違を  
 測定するステップと；

前記減少した数の分析試料を用いて、前記入力データセットのテストデータセットを具  
 える訓練された前記サポートベクタマシンをテストして、最適解が得られたか否かを判定  
 するステップと；

最適解が得られない場合に、前記最適解が得られるまで前記訓練及びテストを繰り返す  
 ステップと；

最適解が得られた直後に、前記コンピュータシステムにフローサイトメトリーデータの実 20

データセットを入力するステップと；

訓練されテストされた前記サポートベクタマシンを用いて前記減少した数の分析試料を用いることによって、前記実データセットを処理して、前記フローサイトメトリデータの分類を具える結果を、2つの区別できる分類のうち的一方に作り出すステップと；

フローサイトメトリデータの分類の識別を具える出力表示を表示装置に作り出すステップと；

を実行することを特徴とする方法。

【請求項2】

前記分散カーネルが、 $p$  及び  $q$  を入力データ点、 $M$  を正規分布の平均値、 $\Sigma$  を共分散行列とすると、

$$k(p, q) = e^{-\rho(p, q)} = \left( \sqrt{\frac{|\Sigma_1 + \Sigma_2|}{2}} \right)^{-1} \exp \left\{ -\frac{1}{8} (M_2 - M_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) \right\}$$

の形式のバッタチャリヤ・アフィニティ (Bhattacharya affinity) を具えることを特徴とする請求項1に記載の方法。

【請求項3】

前記1又はそれ以上の特徴選択アルゴリズムが、リーブワンアウト (leave-one-out) の誤差率分析を具えることを特徴とする請求項1に記載の方法。

【請求項4】

前記1又はそれ以上の特徴選択アルゴリズムが、カーネルアライメントを具えることを特徴とする請求項1に記載の方法。

【請求項5】

前記1又はそれ以上の特徴選択アルゴリズムが、リーブワンアウト (leave-one-out) の誤差率分析及びカーネルアライメントを具える2つの特徴選択アルゴリズムを具えることを特徴とする請求項1に記載の方法。

【請求項6】

前記フローサイトメトリデータが、骨髓異形成症候群をスクリーニング検査される患者から吸引された骨髓サンプルのテスト結果を具えており、

前記フローサイトメトリデータの分類が、前記症候群の存在又は不存在のいずれかを具えることを特徴とする請求項1に記載の方法。

【請求項7】

さらに、前記フローサイトメトリデータが得られた1又はそれ以上のサンプルに対応するデジタル画像データセットをダウンロードするステップと；

画像訓練セット及び画像テストセットに前記デジタル画像データセットを分離するステップと；

前記画像訓練セットを予処理して、前記画像訓練セットの中の関心のある特徴を同定且つ抽出するステップと；

少なくとも1の第2のサポートベクタマシンを訓練して、前記画像訓練セットの中の関心のある特徴を分類するステップと；

前記少なくとも1の第2のサポートベクタマシンをテストして、最適解が得られたか否かを判定し、得られない場合、前記最適解が得られるまで前記訓練及びテストを繰り返すステップと；

前記フローサイトメトリデータが得られた前記実データセットに対応する実画像データを入力するステップと；

前記実画像データを処理して、前記実画像データの中の前記関心のある特徴を分類し、画像の分類を行うステップと；

総合データの分類のために、セカンドレベルのサポートベクタマシンに前記画像の分類及び前記フローサイトメトリデータの分類を入力するステップと；

前記画像の分類及びフローサイトメトリデータの分類の総合結果を具える出力分類を

10

20

30

40

50

作り出すステップと；  
 を具えることを特徴とする請求項 1 に記載の方法。

【請求項 8】

デジタル画像の前記入力データセット及びデジタル画像の前記実データセットが、骨髓異形成症候群をスクリーニング検査される患者の細胞遺伝学的画像であることを特徴とする請求項 7 に記載の方法。

【請求項 9】

前記出力分類が、前記症候群の存在又は不存在のいずれかを具えることを特徴とする請求項 8 に記載の方法。

【請求項 10】

フローサイトメトリデータの分析及び分類のために、コンピュータ可読媒体に内蔵されるコンピュータプログラム製品であって、

前記コンピュータプログラム製品が、サポートベクタマシン分類部を実行し、さらに、  
フローサイトメトリ分析によって複数のサンプルについて実施される複数の分析試料を具える入力データセットを受け取り；

1 又はそれ以上の特徴選択アルゴリズムを具える前記入力データセットの訓練セットを  
 予処理して、前記複数の分析試料の減少した数の分析試料を選択し；

分散カーネルを具える少なくとも 1 の第 1 のサポートベクタマシンを訓練して、前記減少した数の分析試料に対応する前記訓練セットを 2 つの確率分布に分け、前記 2 つの確率分布間の相違を測定するステップと；

前記減少した数の分析試料を用いて、前記入力データセットのテストセットを具える訓練された前記サポートベクタマシンをテストして、最適解が得られたか否かを判定し；

最適解が得られない場合に、前記最適解が得られるまで前記訓練及びテストを繰り返し；

最適解が得られた直後に、前記コンピュータシステムにフローサイトメトリデータの実データセットを受け取り；

訓練されテストされた前記サポートベクタマシンを用いて前記減少した数の分析試料を用いることによって、前記実データセットを処理して、前記実データセットの分類を具える結果を、2 つの区別できる分類のうち的一方に作り出し；

前記実データセットの分類の識別を具える出力表示を表示装置に作り出すことをコンピュータプロセッサに行わせるための命令を具えることを特徴とするコンピュータプログラム。

【請求項 11】

前記分散カーネルが、 $p$  及び  $q$  を入力データ点、 $M$  を正規分布の平均値、 $\Sigma$  を共分散行列とすると、

$$k(p, q) = e^{-\rho(p, q)} = \left( \sqrt{\frac{|\Sigma_1 + \Sigma_2|}{2}} \right)^{-1} \exp \left\{ -\frac{1}{8} (M_2 - M_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) \right\}$$

の形式のバッタチャリヤ・アフィニティ (Bhattacharya affinity) を具えることを特徴とする請求項 10 に記載のコンピュータプログラム。

【請求項 12】

前記 1 又はそれ以上の特徴選択アルゴリズムが、リーブワンアウト (leave-one-out) の誤差率分析を具えることを特徴とする請求項 10 に記載のコンピュータプログラム。

【請求項 13】

前記 1 又はそれ以上の特徴選択アルゴリズムが、カーネルアライメントを具えることを特徴とする請求項 10 に記載のコンピュータプログラム。

【請求項 14】

前記 1 又はそれ以上の特徴選択アルゴリズムが、リーブワンアウト (leave-one-out)

10

20

30

40

50

の誤差率分析及びカーネルアライメントを具える２つの特徴選択アルゴリズムを具えることを特徴とする請求項１０に記載のコンピュータプログラム。

【請求項１５】

前記フローサイトメトリデータが、骨髓異形成症候群をスクリーニング検査される患者から吸引された骨髓サンプルのテスト結果を具えており、

前記結果が、前記症候群の存在又は不存在の表示を具えることを特徴とする請求項１０に記載のコンピュータプログラム。

【請求項１６】

さらに、前記フローサイトメトリデータが得られた１又はそれ以上のサンプルに対応するデジタル画像データセットを受け取り；

画像訓練セット及び画像テストセットに前記デジタル画像データセットを分離し；

前記画像訓練セットを予処理して、前記画像訓練セットの中の関心のある特徴を同定且つ抽出し；

少なくとも１の第２のサポートベクタマシンを訓練して、前記画像訓練セットの中の関心のある特徴を分類し；

前記少なくとも１の第２のサポートベクタマシンをテストして、最適解が得られたか否かを判定し、得られない場合、前記最適解が得られるまで前記訓練及びテストを繰り返す；

前記フローサイトメトリデータが得られた前記実データセットに対応する実画像データセットを受け取り；

前記実画像データを処理して、前記実画像データの中の前記関心のある特徴を分類し、画像の分類を作り出し；

総合データの分類のために、セカンドレベルのサポートベクタマシンに前記画像の分類及び前記フローサイトメトリデータの分類を入力し；

前記画像の分類及びフローサイトメトリデータの分類の総合結果を具える出力分類を作り出すことを前記コンピュータプロセッサに行わせるための命令を具えることを特徴とする請求項１０に記載のコンピュータプログラム。

【請求項１７】

デジタル画像の前記入力データセット及びデジタル画像の前記実データセットが、骨髓異形成症候群をスクリーニング検査される患者の細胞遺伝学的画像であることを特徴とする請求項１６に記載のコンピュータプログラム。

【請求項１８】

前記出力分類が、前記症候群の存在又は不存在のいずれかを具えることを特徴とする請求項１７に記載のコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本出願は、２００８年２月８日に出願された米国仮出願６１／０２７，４１６の優先権を主張するものであり、この出願の開示はここにすべて引用されている。

【０００２】

本発明は、サポートベクタマシンを用いた、分散データ、特にフローサイトメトリデータの自動分析のための方法に関する。

【背景技術】

【０００３】

フローサイトメトリーは、流動液体流に懸濁する微小粒子の特性の測定である。レーザー光の収束ビームが、運動する各粒子を照射し、光があらゆる方向に散乱する。交差点の前方に又はレーザービームに直交するよう配置される検出器が、散乱光のパルスを受信し、分析のためのコンピュータ分析器の中に入力される信号を発生する。検出される前方散乱光の総量は、粒径及び屈折率に依存するが、レーザーの側から見る粒子の断面積に密接に関連し、側方散乱光の量は形状又は粒度を示し得る。

10

20

30

40

50

## 【 0 0 0 4 】

フローサイトメトリーの最も広範に使用される適用例の1つは、医療診断のための細胞レベルの分析であり、関心のある粒子は食塩水を含む溶液に懸濁する細胞である。また、関心のある細胞マーカを蛍光染料によってラベル付けし得るならば、表面分子又は細胞間構成といった細胞のさらなる特性を正確に定量化し得る；例えば、抗体蛍光染料複合体を用いて特定の表面又は受容体に付着し得る。表面マーカに対する蛍光標識モノクローナル抗体の使用を通して様々な進行段階で細胞を特徴付けることによる免疫表現型検査は、フローサイトメトリーの最も一般的な適用例の1つである。特定の構造体（例えば、DNA、ミトコンドリア）に結合する、又は局所化学（例えば、 $Ca^{++}$ の濃度、pH、等）に敏感な他の染料が開発されている。

10

## 【 0 0 0 5 】

フローサイトメトリーが医療診断に広く使用される一方、水又は他の液体の分析といった医療以外での適用にも有用である。例えば、海水を分析して微生物又は他の生物の存在又はそのタイプを特定することができ、細菌を検査するために牛乳を分析することができ、粒子状の不純物又は添加剤について燃料を検査し得る。

## 【 0 0 0 6 】

使用されるレーザービームは、選択される蛍光色素又は蛍光色素を励起するのに適切な色である。発光する蛍光の光量は、問題となっている細胞マーカの発現に関連する。各フローサイトメトリーは、通常、その構成に応じて多くの様々な蛍光色素を同時に検出し得る。ある機器では、様々な波長で発射される複数のレーザーを使用することによって、複数の蛍光色素を同時に分析し得る。例えば、Becton Dickinson (Franklin Lakes, NJ)から市販されているFACSCalibur™ フローサイトメトリーシステムは、4色オペレーションのために構成される多色フローサイトメータである。各細胞からの蛍光放射が集束され、蛍光強度値をFlow Cytometry Standard (FCS) データファイルの各信号に割り当てるコンピュータで分析される。データ分析は、さらなる分析又は分類のために、データをフィルタリング又は「ゲーティング」するのに使用され事象の部分母集団の一部を規定する超空間の多角形領域の交差又は結合を特定することを含む。

20

## 【 0 0 0 7 】

International Society for Analytical Cytology (ISAC)は、FCSデータの共通表現のためにFCSデータファイル基準を採用している。この基準は、全ての主要な分析機器によってサポートされており、サイトメータを通過するサンプルからの測定値を記録し、研究者及び臨床医は、多くのデータ適合性の問題に直面せずに、多くの市販の機器及びソフトウェアの中から選択し得る。しかしながら、このような基準は、コンピュータによる予処理及びデータ分析のためのプロトコルを記述するまでに至っていない。

30

## 【 0 0 0 8 】

フローサイトメトリー分析に存在する大量のデータのため、多くの場合、手作業によりデータを完全に使用することは困難である。また、高次元のデータが、従来の統計的方法及び人工の神経ネットワークといった学習方法を使用することを不可能にしている。サポートベクタマシンは、高次元のデータを処理し得る方法を学習するカーネルベースのマシンである。それは、適切に設計されたカーネルを具えた流量データを取り扱う際に効果的なツールである。

40

## 【 0 0 0 9 】

カーネルは、サポートベクタマシン (SVM) といった現状のマシン学習方法で重要な役割を果たす。分類のためのサポートベクタマシンは、多くの場合、高次元（無限次元でもよい）の内積空間である特徴空間での最適な超平面で定義される。最適な超平面の構造は、特徴空間におけるマッピングされた入力ベクトルの内積を要する。入力空間で定義されるカーネル関数は、特徴空間に入力を実際にマッピングせずに、内積を計算するための効果的な方法を与える。カーネルは、2つのベクトル間の類似性測度を定義する。SVM法の利点は、明確な特徴抽出を用いずに、単にカーネルの通常の類似性測度に基づいて隠

50

れたパターンを採取するその能力である。

【 0 0 1 0 】

画像認識及びフローサイトメトリデータ分析といった多くの適用例において、入力データは、通常高次元であり大量である。このようなデータの重要な特徴は、通常、個々のポイントが孤立した値ではなくて、特定の空間においてポイントが分散している。標準的カーネルは、多くの場合、大きな入力ボリュームがカーネルを基底構造及び特定の問題の分布特徴に対して敏感にさせないように、全てのベクトル成分を等しく処理するため、このようなタイプのデータに役に立たない。結果として、それらは、分散データにあまり適さない。例えば、フローサイトメトリデータの SVM 分析が、ガウスカーネル及び B - スプラインカーネルが例として挙げられる放射基底関数 ( R B F ) カーネルを用いて報告されている。(Rajwa, B.らの「Automated Classification of Bacterial Particles in Flow by Multiangle Scatter Measurement and Support Vector Machine Classifier」 Cytometry Part A, 73A:369-379 (2008)を参照。)説明されている方法は、報告された分類の高度な正確さを得るために「強化された散乱検出システム」の使用を要した。さらに、著者は、SVMの結果は、問題の次元が2よりも高いか否かを容易に判断し得ないと結論付けた。このような制限は、このような方法の実際の適用を最小限にする。Toedlingらは、「Automated in-silico detection of cell populations in flow cytometry readouts and its application to leukemia disease monitoring」, BMC Bioinformatics, 7:282, June 2006,で、単細胞を予め規定されたグループに割り当てることによって白血病を検出するために、放射基底関数カーネルを用いたフローサイトメトリデータの SVM 分析を説明している。実際には、SVM分析は、手動のゲーティングの代わりにするが、データの分散特徴を考慮に入れない。

【 0 0 1 1 】

したがって、フローサイトメトリデータ及び基底構造及び分散の中に重要な情報を含む他のタイプの分散データを分析するための、より高次元で使用する必要性がある。本発明は、このような方法を対象としている。

【発明の概要】

【 0 0 1 2 】

本発明によれば、分散特性を有するデータ用に特別に形成された SVM カーネルを用いて、フローサイトメトリデータを分析する。入力データ p セットは、空間中の多数のポイントの集まりである。例えば、画像を2次元空間のポイントのセットと見なすことができる。適切な規格化の後に、p を確立分布と見なすことができる。2つのこのような入力データ p 及び q のカーネルを定義して分散傾向を得るために、p 及び q に関する関数を定義して、分散中の単なる個々のポイントではなく、2つの全体的な分散間の類似度を測定する。

【 0 0 1 3 】

例として、特にここで引用される米国特許第 6, 760, 715 号、第 7, 117, 188 号及び第 6, 996, 549 号に概略的に開示されたサポートベクタマシンを使用して、市販されている従来のフローサイトメトリセットアップによって生成されるフローサイトメトリデータを分析する。フローサイトメトリ測定を実行するための典型的なシステムは、ここで引用される米国特許第 5, 872, 627 号及び第 4, 284, 412 号に記載されている。ここで記載されている特別な実施例では、データは、特に骨髓異形成症候群 ( M D S ) といった血液学的状態を検査するための医療診断での適用に関する。フローサイトメトリの免疫表現型検査は、形態学及び細胞学の組み合わせが非診断的である場合でさえも、造血細胞の量的且つ質的異常を検出するための正確且つ高感度な方法であることが実証されている。

【 0 0 1 4 】

本発明によれば、フローサイトメトリデータの入力を受け取り、1又はそれ以上のサポートベクタマシンを用いてこのデータを分析し、フローサイトメトリデータが2又はそれ以上のカテゴリに分類される出力を出すための自動化方法及びシステムが提供される

。1又はそれ以上のサポートベクタマシンは、入力データの中の分散データを取得するカーネルを使用する。このような分散カーネルは、2つの分散間の距離関数（発散）を使用することによって構築される。2つの確率分布間の相違を測定する適切な分散関数の例は、Kullback-Leibler発散、バッタチャリヤ・アフィニティ（Bhattacharya affinity）、Jeffrey発散、Mahalanobis距離、Kolmogorov 変分距離、及び予測条件付きエントロピを含んでいる。好適な実施例では、バッタチャリヤ・アフィニティを使用する。分散カーネルをSVMに直接又は他のマシンに適用して、分類器又は他の予測システムを形成する。

【図面の簡単な説明】

【0015】

【図1】図1は、本発明に係るフローサイトメトリデータの自動収集及び分析のためのシステムの概略図である。

10

【図2】図2は、MDSのフローサイトメトリ分析における関心のある集団の分布の典型的な両対数表示である。

【図3】図3は、本発明に係るデータ分析法のフローチャートである。

【図4】図4は、MDSテストのフローサイトメトリデータの分析に対する独創的な方法の適用から得られる受信者動作特性（ROC）曲線を示すプロットである。

【発明を実施するための形態】

【0016】

本発明によれば、フローサイトメトリデータの分析のための方法及びシステムが提供される。特に、このような独創的な方法が、分散特性のデータ分析に使用するためのカーネルの生成を含む。フローサイトメトリアプリケーションの入力データpは、空間内の多くのポイントの集りである。例えば、画像を2次元空間におけるポイントのセットと見なすことができる。適切な規格化の後に、pを確率分布と見なすことができる。このようなp及びqのカーネルを定義して分散傾向を取得するために、分散の個々のポイントではなく2つの全体的な分散間の類似度を測定するp及びqに関する関数を定義する必要がある。

20

【0017】

このような「分散カーネル」を構築するための1つの方法は、2つの分散間の距離関数（発散）を使用することである。p（p，q）を距離関数とすると、カーネルは以下のとおりである。

$$k(p, q) = e^{-\rho(p, q)}$$

30

2つの確率分布間の相違を測定する多くの距離関数が存在する。Kullback-Leibler発散、Bhattacharya affinity、Jeffrey発散、Mahalanobis距離、Kolmogorov 変分距離、及び予測条件付きエントロピは、全てこれらの距離関数の例である。距離関数を与えることによって、上記の式を用いてカーネルを構築し得る。例えば、Bhattacharya affinityに基づいて、特別なカーネルを構築し得る。平均値M及び共分散行列Σを具えた正規分布では、Bhattacharya affinityは、式：

$$\rho(p, q) = \frac{1}{8} (M_2 - M_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) + \frac{1}{2} \ln \frac{|\Sigma_1 + \Sigma_2|/2}{\sqrt{|\Sigma_1| \cdot |\Sigma_2|}}$$

を有する。

40

このような距離関数から、新たなカーネルが上記の式を用いて定義される。

$$k(p, q) = e^{-\rho(p, q)} = \left( \frac{|\Sigma_1 + \Sigma_2|/2}{\sqrt{|\Sigma_1| \cdot |\Sigma_2|}} \right)^{-1} \exp \left\{ -\frac{1}{8} (M_2 - M_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) \right\}$$

このような分散カーネルは、線形複雑度を具えて計算上効果的であり、大量の入力データを扱い得る。典型的な密度推定法が、ある適用例では高過ぎる計算の複雑性O(n<sup>2</sup>)を有する。独創的な分散カーネルをSVM又は他のマシン学習システムに直接的に適用して、分類器又は他の予測システムを形成し得る。分散カーネルは、SVM及び他のカーネルマシンで頻繁に使用される標準カーネルに対していくつかの際だった利点を与える。それ

50

らは、ある適用例では重要な大きなデータ成分の全ての分散間の類似度を取得する。

【0018】

図3は、フローサイトメトリーデータの分析に使用される典型的な処理の流れを示す。当業者にとって容易に明らかなように、フローサイトメトリーデータは分散データの一例として与えられ、以下に説明する方法を用いて、他のタイプの分散データを処理及び分類し得る。

【0019】

フローサイトメータ106によって生成される実データは、1又はそれ以上のサポートベクターマシンを実行するようプログラムされた少なくともメモリ及びプロセッサを含むコンピュータ処理システムに入力される(ステップ302)。一般的なパーソナルコンピュータ(PC)又はAPPLE(登録商標)MACINTOSH(登録商標)タイプのプロセッサが、このような処理に適している。セットされる入力データは2つの部分に分割され、一方はサポートベクターマシンを訓練する際に使用され、他方は訓練の有効性をテストする際に使用される。ステップ304で、特徴選択アルゴリズムが、プロセッサの中で1又はそれ以上の特徴選択プログラムを実行することによって、訓練データセットで動作する。ステップ306で、特徴を減らした訓練データセットが、バッタチャリヤ・アフィニティー(Bhattacharya affinity)ベースのカーネルといった分散カーネルを具えたサポートベクターマシンを用いて処理される。独立したテストデータセットの中からステップ304で選択された特徴に対応するデータを抽出し、分散カーネルを具えた訓練されたSVMを用いてテストデータを処理することによって、ステップ308で訓練ステップの有効性が見積もられる。テストの結果が最適な結果よりも低いことを示す場合、最適解が得られるまでSVMを再び訓練して再テストを行う。訓練が満足がいくものと判断される場合、患者試料から取り込まれるフローサイトメトリー測定値に対応する実データが、ステップ310でプロセッサに入力される。ステップ304で選択された特徴は患者試料から選択され、ステップ312で、分散カーネルを具えた訓練且つテストされたSVMを通して処理され、結果は正常又は異常として患者試料を分類する。ステップ314で、結果を要約する報告が作成され、それがコンピュータのモニタ122、プリントされる結果124に表示され、及び/又はeメール又は他のネットワークファイル伝達システムを介して、研究又は医療機関、病院又は診療所に送信される。また、データグループの1次元及び2次元表示を具えたヒストグラムが、表示及び/又は印刷される。また、実データ、ヒストグラム及び他の患者データとともに、コンピュータのメモリ又は患者のデータベースに結果が記憶される。

【0020】

最適な付加的診断方法をフローサイトメトリーデータ及び結果と組み合わせて、自動分析システムの信頼性を高めることができる。ここに引用されているZhangらの米国特許第7,383,237号に開示されたのと同様な方法を使用して、フローサイトメトリーテストの結果を、他のタイプのテストと組み合わせることができる。図3は、染色体分析又は蛍光インシトゥハイブリデーション法(FISH)といった従来の処理で生成される染色体の画像から、関心のある特徴を抽出することによって、SVMを用いたコンピューター支援による細胞遺伝学データの画像解析を実行し、欠失、転座、逆位及び他の異常を特定するためのフローチャートを示す。ステップ320で、訓練用画像データが、コンピュータのプロセッサに入力され処理されて、関心のある特徴を特定且つ抽出する。一般に、訓練用画像データは、予処理されて関心のある特徴を特定し(ステップ322)、その後画像処理SVMを訓練するよう使用される。そして、テスト用画像データを使用して最適解が得られたかを検証する(ステップ324)。得られない場合、ステップ324を繰り返し、SVMを再訓練して再テストする。最適解が得られた場合、患者の実画像データが入力され(ステップ326)、予処理(ステップ328)及び分類(ステップ330)される。

【0021】

好適な方法では、特許第7,383,237号に記載されているように、画像の中の関

10

20

30

40

50

心のある各特徴が、そのような特徴を最適化するSVMによって別々に予処理（ステップ322）及び処理される。関心のある全ての特徴の分析結果が、セカンドレベルの画像処理SVMで組み合わせたり、画像全体を分類する出力を生成する。訓練されたSVMは、予処理された画像テストデータを用いてテストされる（ステップ324）。解が最適である場合、患者（フローサイトメトリー分析を実行する同じ患者）の実データに対応する画像がプロセッサに入力される（ステップ326）。患者の画像データが予処理されて関心のある特徴を特定し（ステップ328）、関心のある各特徴が、特定の特徴のために最適化される訓練された第1のレベルSVMを通して処理される。関心のある特徴の分析の総合結果が組み合わせられ、訓練されたセカンドレベルの画像処理SVMに入力され、画像全体を分類する出力を生成する（ステップ330）。

10

#### 【0022】

ステップ330の結果を、患者のデータベースの患者ファイルに保存のために送ることができ（ステップ316）、ステップ312から得られるフローサイトメトリーデータと総合分析のためにセカンドレベルのSVMに入力される。セカンドレベルSVMは、ステップ308、324及び340間の点線で示すように訓練及びテストデータを用いてすでに訓練且つテストされている。ステップ316及びステップ330の結果が、ステップ342で、総合分析のために訓練されたセカンドレベルSVMの処理のために組み合わせる。このような総合処理の結果は、一般的に、例えば正常又は異常、罹患又は非罹患といったバイナリ出力である。ステップ314で表示のために総合結果を出力し、及び/又は保存のためにメモリ又は患者のデータベースに入力する（ステップ316）。さらなる任意の第2のフロー経路が、専門家による分析、患者の履歴等といった、他のタイプのデータ及び分析を組み込むために提供され、スクリーニング、モニタリング及び/又は治療に使用する最終的診断又は予後スコア又は他の出力を生成するよう組み合わせる。

20

#### 【0023】

実施例1：骨髓異形成症候群（MDS）の検出

本研究の目的は、細胞遺伝学の染色体異常に関連する骨髓異形成症候群（MDS）とフローサイトメトリーデータのパターンとの潜在的な関係性を調査することである。この免疫表現型分析は、フローサイトメトリーの最も一般的な適用例のうちの1つであり、試料採取及び調整の手順が当業者によく知られている。図1に示す流れにしたがって、MDSを患う疑いのある患者からの骨髓穿刺液102が、生理食塩水、又はヘパリンナトリウム溶液に採取され、サイトメータシステム106のフローセルの中に懸濁液を導入するよう構成された多数の試験管104又は他の容器に細胞懸濁液を作製する。様々な蛍光色素に抱合されるモノクローナル抗体を含む試薬が試験管の中に導入され、それぞれの試験管は、様々な抗体の組み合わせであって、いくつかの可能性のある蛍光色素のうちの1つに抱合される様々な抗体の組み合わせを受け入れる。フローサイトメータは、Becton Dickinson (Franklin Lakes, NJ)のFACSCalibur™又はOrtho Diagnostics (Raritan, NJ)のthe Cytoron/ Absolute™を含む多くの製造者から市販されている。例えば、4色測定のために、FACSCalibur™システムを使用した。当業者に明らかなように、このようなシステムは、示されるものがフローサイトメータの分析フィールドの中の試料の存在のみを示す図表を意図するように、カルーセルの中に装填される複数の試料の自動操作を提供する。フローサイトメータシステム106の前方散乱検出器108及び側方散乱検出器110が、分析の流れを通して細胞が検出される際に、検出された事象に対応する電気信号を発生させる。側方散乱検出器110の中に含まれる蛍光検出器が、様々な蛍光マーカーに抱合される抗体によって示されるように、抗原の発現によって発生する蛍光信号の振幅を測定する。多くの値が、様々な検出器のそれぞれによって測定されるパルスの高さ（振幅）に基づいて生成される。得られる信号が、コンピュータワークステーションの120の中のプロセッサの中に入力され、グラフ表示モニタ122での表示のために、検出事象に対応するヒストグラム（1つ又は2つのパラメータ）を作成するよう使用される。比較参照試料との比較に基づいた正常又は異常に関する入力データの分類を含む、本発明に係るこのようなデータの分析により、プリントされ又はモニタ122上に表示され得るレポート12

30

40

50

4を得る。また、実データ、ヒストグラム及びレポートが、患者の他の記録と関連させるために、コンピュータワークステーション120の内部メモリ及び医療機関又は他の医療施設のデータウェアハウスの一部であるデータベースサーバ130を含む別個の記憶装置の一方又は双方に保存される。

#### 【0024】

入力データセットは、フローサイトメトリー及び細胞遺伝学データの双方を有する77のケース（患者）を含んでいる。全ての患者は、MDSを患っている疑いがある。数値又は構造の変化に関する染色体全体の顕微鏡検査を含む細胞遺伝学的試験が示すように、77のケースのうち、37が染色体異常を有した。残りの40は、細胞遺伝学的には陰性であることが分かった。

10

#### 【0025】

懸濁液の中の吸引された骨髓試料は、各患者について13の試験管に分けられた。標準的な4色の免疫蛍光法の手順では、前方光散乱（FSC）及び直交光散乱（SSC）が4色抗体の組み合わせとともに集められて7つの異なる分析を行ったが、そのうちの1つはブランクであった。各ケースは、一般に、全ての分析試料を検査する場合、20,000乃至50,000の事象を有する。各ケースについて得られるフローサイトメトリーデータセットは、約 $10^6$ の測定値を有した。図2は、マークされた様々な細胞集団の側方散乱対CD45表現を示す典型的なヒストグラムを示す。

#### 【0026】

13の試験管のそれぞれについて、FSC及びSSCを測定し、図2の左下欄に示すように、ゲーティング（gating）が細胞残屑を排除し得た。さらに、蛍光マーカを具えた抗原特異性の様々な組み合わせを各試験管について用いた。表1は、以下のマーカ：FITC（fluorescein イソチオシアネート）、PE（フィコエリトリン）、PerCP（peridinin-クロロフィル）、及びAPC（アロフィコシアニン）を具えたモノクローナル抗体の様々な組み合わせをまとめるものである。特定の蛍光マーカに抱合されるモノクローナル抗体は、Becton-Dickinson Immunocytometry Systems (San Jose, CA)、DakoCytomation (Carpinteria, CA)、Caltag (Burlingame, CA)及びInvitrogen Corporation (Camarillo, CA)を含む多くの様々な供給元から市販されている。成熟リンパ球の計数に使用されるCD45抗体は、リンパ球ゲーティングの検証のための各組み合わせに含まれる。

20

30

表 1

| 試験管 | FITC- 抱合           | PE- 抱合                | PerCP- 抱合      | APC- 抱合          |
|-----|--------------------|-----------------------|----------------|------------------|
| 1   | IgG1               | IgG1 +PI              | CD45/2D1/IgG1  | IgG1             |
| 2   | IgG2b              | IgG2b                 | CD45/2D1/ IgG1 | IgG2b            |
| 3   | CD8/SK1/IgG1       | CD2/RPA-2.10/IgG1     | CD45/2D1/IgG1  | CD4/SK3/IgG1     |
| 4   | CD7/M-T701/IgG1    | CD56/MY31/IgG1        | CD45/2D1/IgG1  | CD3/SK7/IgG1     |
| 5   | CD19/SJ25C1/IgG1   | CD23/M-L233/IgG1      | CD45/2D1/IgG1  | CD5/UCHT-2/IgG1  |
| 6   | CD22/S-HCL-1/IgG2b | CD10/HI10a/IgG1       | CD45/2D1/IgG1  | CD34/8G12/IgG1   |
| 7   | CD10/HI10a/IgG1    | CD11c/S-HCL-3/IgG2b   | CD45/2D1/IgG1  | CD20/2H7/IgG2b   |
| 8   | CD38/HB.7/IgG1     | Dako Kappa/F(ab)2rab  | CD45/2D1/IgG1  | CD20/2H7/IgG2b   |
| 9   | CD38/HB.7/IgG1     | Dako Lambda/F(ab)2rab | CD45/2D1/IgG1  | CD20/2H7/IgG2b   |
| 10  | Kappa Caltag poly  | Lambda Caltag poly    | CD45/2D1/IgG1  | CD19/SJ25C1/IgG1 |
| 11  | HLA-DR/TU36/IgG2b  | CD117/104D2           | CD45/2D1/IgG1  | CD11b/Mac-1/IgG1 |
| 12  | CD14/MoP9/IgG2b    | CD13/L138/IgG1        | CD45/2D1/IgG1  | CD64/10.1/IgG1   |
| 13  | CD16/NKP15/IgG1    | CD33/P67.6/IgG1       | CD45/2D1/IgG1  | CD34/8G12/IgG1   |

40

SVMの訓練及び訓練の評価双方のためのデータを与えるために、77ケースのデータセット全体を訓練セット及び独立したテストセットに分割した。40のケース（細胞遺伝学

50

的試験によって判定されるように20が陽性で20が陰性)を使用して、SVMを訓練した。残りの37のケース(17が陽性で20が陰性)を使用して、独立したテストセットを形成した。

【0027】

フローサイトメトリーデータの分析にバッチチャリヤアフィニティに基づく上述の特別のカーネルを使用して、2つの確率分布間の相違を測定した。

【0028】

分類器の全ての分析試料からのデータの介入は、最適性能を具えたシステムを生成しない。このため、分析試料の特徴選択が訓練用セットに基づいて行われる。2つの性能試験を特徴選択ステップに適用した。第1の特徴選択法であるSVMに関するleave-one-out(LOO)誤差率は、初期のデータセットについてのSVMを訓練し、その後で、LOOエラーが減少するように傾斜ステップを実行することによってスケールリングパラメータを更新することを含む。これらにステップを、LOOエラーの最小に達するまで繰り返す。停止基準を適用し得る。第2の特徴選択法は、カーネルアライメントであった。このような方法は、ここに引用されているCristianiniの米国特許7,299,213に説明されている。カーネルアライメントは、訓練用データをのみ使用して、カーネルマシンの訓練が発生する前に実行し得る。

【0029】

特徴選択プロセスの間に、かなりの数の特徴が正確なデータ分類に寄与することを判断した。特徴選択手順の結果を表2に示す。

表2

| 試験管 # | 分析試料 | ブランク | FSC | SSC | マーカ  |    |       |     |
|-------|------|------|-----|-----|------|----|-------|-----|
|       |      |      |     |     | FTIC | PE | PerCP | APC |
| 1     | 0    | 1    | 0   | 0   | 0    | 1  | 0     | 0   |
| 2     | 0    | 0    | 1   | 0   | 0    | 0  | 0     | 1   |
| 3     | 0    | 1    | 1   | 0   | 0    | 0  | 0     | 0   |
| 4     | 0    | 0    | 1   | 1   | 0    | 0  | 0     | 1   |
| 5     | 0    | 0    | 0   | 0   | 0    | 0  | 0     | 1   |
| 6     | 0    | 0    | 1   | 0   | 0    | 0  | 0     | 0   |
| 7     | 0    | 1    | 1   | 1   | 0    | 0  | 0     | 0   |
| 8     | 0    | 1    | 1   | 1   | 0    | 0  | 0     | 1   |
| 9     | 0    | 1    | 1   | 1   | 0    | 0  | 0     | 0   |
| 10    | 0    | 0    | 1   | 1   | 0    | 0  | 0     | 0   |
| 11    | 0    | 1    | 1   | 0   | 0    | 0  | 0     | 1   |
| 12    | 0    | 0    | 0   | 0   | 0    | 0  | 0     | 0   |
| 13    | 0    | 0    | 0   | 0   | 0    | 0  | 0     | 0   |

表2の入力値「1」は、特定の分析試料(試験管/分析試料の組み合わせ)が選択されることを示し;「0」は分析試料を選択しなかったことを示す。これは、データを分類するために各ケースから考えられる特徴の数を、元の91から26に減らす。そして、減少した分析試料の数からのデータを使用して、分散カーネルを具えたSVMを訓練した。

【0030】

その後、選択した分析試料を用いて、37の独立したケースについて、訓練したSVMをテストした。バイナリ分類テストを実行する従来の統計的測定を用いて、0のカットオフでの結果をまとめた。感度、又はリコールレートが、細胞遺伝学的テストによって判定されるように、陽性の総数に対する正確に分類された陽性の割合を与える。特異度が、正確に特定される陰性の割合を測定する。検査データの分析結果は、以下のようであった:

感度：  $15 / 17 = 88\%$       特異度：  $19 / 20 = 95\%$

【0031】

これは、  $3 / 37 = 8\%$  の全体の誤差率を与える。二項分布に関する予測された標準偏差、  $\sigma = 0.0449$  を用いて、テストは、誤差率が  $15\%$  よりも低い  $95\%$  の信頼度を与えた。

【0032】

感度と特異度との間の兼ね合いが、図4に示す受信者動作特性（ROC）曲線にプロットされている。曲線下面積（AUC）は、  $0.974$  であった。

【0033】

上述の分析結果は、細胞遺伝学的試験によって示されるフローサイトメトリのパターンと特定の染色体異常との間の顕著な関係を呈した。

10

【0034】

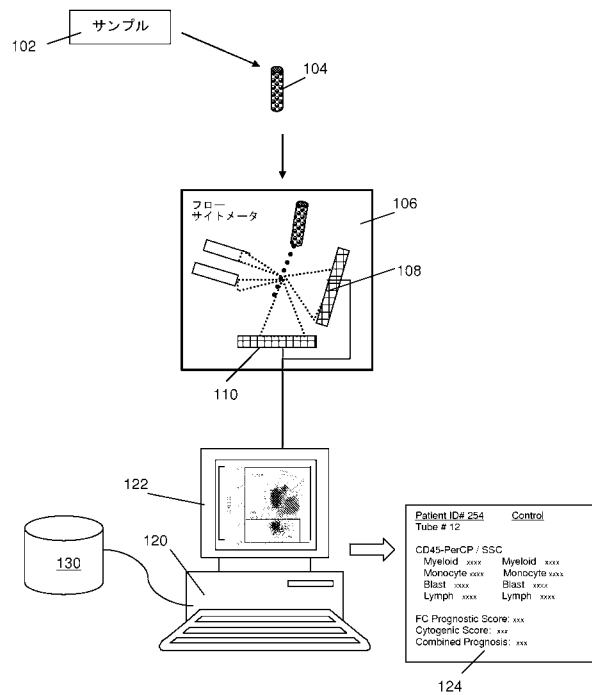
フローサイトメトリデータの複雑さのために、必要な特徴を明確に抽出し又は細胞遺伝学的結果を予測するパターンを規定することは困難である。SVMベースのシステムは、実施例の間の類似性測度のみを要するという際立った優位性を与え、分類器を構成する。

【0035】

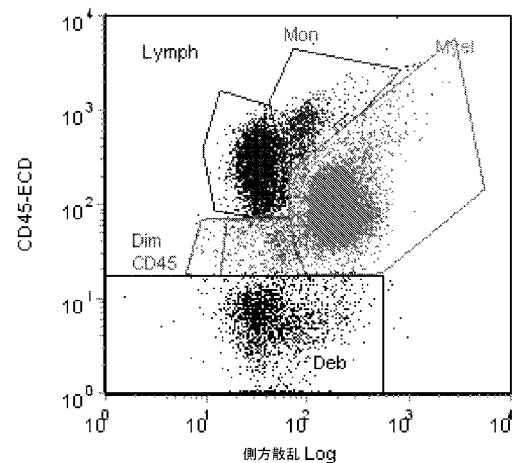
本発明の特別設計されたカーネルは、フローサイトメトリデータの分析で効果を発揮する。このようなカーネルは、このような問題に効果的な計算の複雑性  $O(n)$  を有する。SVM分析の結果は、カーネルもまた非常に有効であることを示す。

20

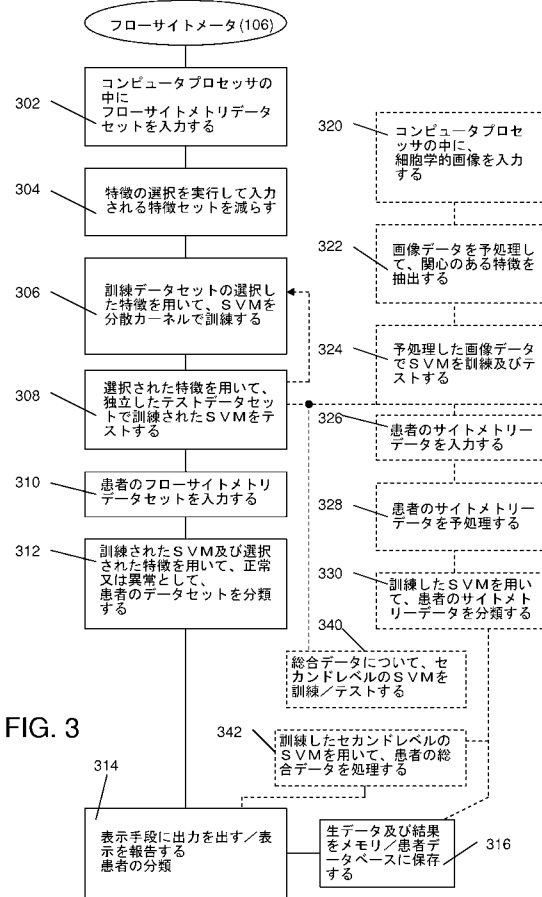
【図1】



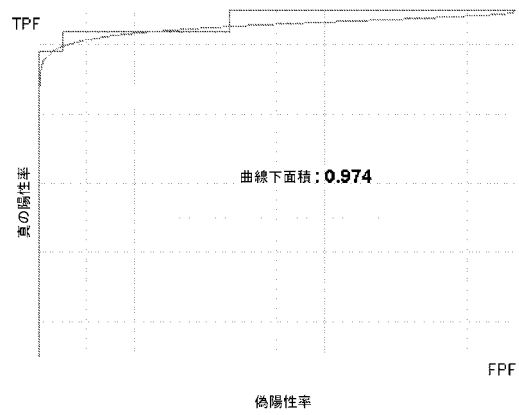
【図2】



【図 3】



【図 4】



---

 フロントページの続き

(51)Int.Cl. F I  
G 0 1 N 21/64 F

(72)発明者 チャン, ホン  
アメリカ合衆国 ジョージア州 3 1 4 4 1, サバナ, ブラックホークトレイル 2 2

審査官 吉田 将志

(56)参考文献 特表 2 0 0 3 - 5 2 9 1 3 1 ( J P , A )  
特表 2 0 0 3 - 5 0 0 7 6 6 ( J P , A )  
特表 2 0 0 7 - 5 0 8 8 1 2 ( J P , A )  
国際公開第 2 0 0 6 / 0 8 4 2 7 2 ( WO , A 1 )  
Joern Toedling , Automated in-silico detection of cell populations in flow cytometry readouts and its application to leukemia disease monitoring , BMC Bioinformatics , 2 0 0 6 年 , Vol.7/Article.282 , 1-11  
Tony Jebara , Probability Product Kernels , Journal of Machine Learning Research , 2 0 0 4 年 , Vol.5 , 819-844  
Risi Kondor , A Kernel Between Sets of Vectors , Proceedings of the Twentieth International Conference on Machine Learning , 2 0 0 3 年 , U R L , [http://machinelearning.wustl.edu/mlpapers/paper\\_files/icml2003\\_KondorJ03.pdf](http://machinelearning.wustl.edu/mlpapers/paper_files/icml2003_KondorJ03.pdf)

(58)調査した分野(Int.Cl. , D B 名)  
G 0 1 N 3 3 / 4 8 - 9 8  
G 0 1 N 1 5 / 1 4  
G 0 1 N 2 1 / 1 7  
G 0 1 N 2 1 / 6 4